Some results on anticoncentration

Sanjana Das

May 12, 2024

1 Introduction

Given a sequence $\mathbf{v} = (v_1, \dots, v_n)$ of n real numbers, we can consider a random signed sum of $v_1, \dots, v_n$ — we imagine choosing $\varepsilon_1, \dots, \varepsilon_n \in \{-1, 1\}$ uniformly and independently at random, and we consider the random variable $\varepsilon_1 v_1 + \cdots + \varepsilon_n v_n$. We are interested in how concentrated this random variable can be at a single point; we formalize this using the following definition.

Definition 1.1. For $\mathbf{v} = (v_1, \dots, v_n) \in \mathbb{R}^n$, we define the *concentration probability* $p(\mathbf{v})$ as $p(\mathbf{v}) = \max_{a \in \mathbb{R}} \mathbb{P}[\varepsilon_1 v_1 + \cdots + \varepsilon_n v_n = a]$ for independent and uniform $\varepsilon_1, \dots, \varepsilon_n \in \{-1, 1\}$.

The problem of finding upper bounds on $p(\mathbf{v})$ was first considered by Littlewood and Offord in [5], who proved that for any nonzero $v_1, \dots, v_n$ we have $p(\mathbf{v}) = O(n^{-1/2} \log n)$. This bound was later improved by Erdős [2], who removed the factor of $\log n$ — Erdős proved a bound of

$$p(\mathbf{v}) \le \frac{\binom{n}{\lfloor n/2 \rfloor}}{2^{-n}} = O(n^{-1/2}). \tag{1}$$

This bound is tight — taking $\mathbf{v} = (1, 1, \dots, 1)$ achieves equality. But we can then ask whether imposing certain restrictions on $v_1, \dots, v_n$ — such as requiring them to be distinct — results in a better bound. This was first considered by Erdős and Moser [1], who proved that if $v_1, \dots, v_n$ are distinct, then $p(\mathbf{v}) = O(n^{-3/2}(\log n)^{3/2})$; Sárközy and Szemerédi [6] improved this bound to

$$p(\mathbf{v}) = O(n^{-3/2}). \tag{2}$$

This bound is tight as well — taking $\mathbf{v} = (1, 2, \dots, n)$ achieves equality. (To see this, note that $\mathrm{Var}[\sum \varepsilon_i v_i] \le n^3$, so by Chebyshev's inequality $\mathbb{P}[|\sum \varepsilon_i v_i| \le 2n^{3/2}] \ge \frac{3}{4}$, which means there is some $a \in [-2n^{3/2}, 2n^{3/2}]$ for which $\mathbb{P}[\sum \varepsilon_i v_i = a] \ge \frac{3}{16} n^{-3/2}$.)

We can then ask whether imposing stronger restrictions on $\mathbf{v}$ lets us improve the exponent even further. In particular, the condition that $v_1, \dots, v_n$ are distinct can be written as $v_i - v_j \ne 0$ for all $i \ne j$; we can ask whether forbidding larger linear relations allows us to get a stronger bound. The answer is yes, as proved by Halász [3].

Theorem 1.2 (Halász 1977)**.** *Given $r \in \mathbb{N}$ and $\mathbf{v} = (v_1, \dots, v_n)$, let R be the number of relations of the form $\xi_1 v_{i_1} + \cdots + \xi_{2r} v_{i_{2r}} = 0$ satisfied by $\mathbf{v}$, over all choices of signs $\xi_1, \dots, \xi_{2r} \in \{-1, 1\}$ and indices $i_1, \dots, i_{2r} \in [n]$. Then $p(\mathbf{v}) = O_r(Rn^{-2r-1/2})$.*

We think of r as fixed and n as large. Note that there are $\Theta_r(n^r)$ 'trivial' relations of the form $\xi_1 v_{i_1} + \cdots + \xi_{2r} v_{i_{2r}} = 0$ — i.e., relations that are identically true, such as $v_1 - v_2 - v_1 + v_3 + v_2 - v_3 = 0$. So Halász's theorem implies that if $\mathbf{v}$ does not satisfy any 'nontrivial' relations of this form, then $p(\mathbf{v}) = O_r(n^{-r-1/2})$.

The results described so far provide upper bounds on $p(\mathbf{v})$ given certain conditions on $\mathbf{v}$. Tao and Vu [8] approach the problem of anticoncentration from a different angle, that of finding inverse theorems — if we

know $p(\mathbf{v})$ is 'large,' then what can we say about $\mathbf{v}$? (All the upper bounds above are of the form $1/\mathsf{poly}(n)$; in contrast, if $v_1, \ldots, v_n$ are completely generic, then $p(\mathbf{v}) = 2^{-n}$ is exponentially small. So we think of $p(\mathbf{v})$ as 'large' if it is $1/\mathsf{poly}(n)$.)

More specifically, we would like to say that if $p(\mathbf{v})$ is large, then $\mathbf{v}$ has a strong 'additive structure' in some sense. (Halász's theorem implies that $\mathbf{v}$ must satisfy many small linear relations, but this alone isn't enough to make $p(\mathbf{v})$ large — for example, if half of $v_1, \ldots, v_n$ have a lot of additive structure but the other half are completely generic, then $p(\mathbf{v})$ will still be exponentially small. So we would really like a statement saying that $\mathbf{v}$ possesses a lot of 'global' additive structure, rather than just 'local' structure.)

As motivation, here is a class of examples for which $p(\mathbf{v})$ is large.

Definition 1.3. For $\mathbf{w} = (w_1, \ldots, w_d) \in \mathbb{R}^d$ and $k \in \mathbb{N}$, we use $\mathbf{Q}(\mathbf{w}, k)$ to denote the set

$$\mathbf{Q}(\mathbf{w}, k) = \{a_1 w_1 + \cdots + a_d w_d \mid a_i \in \{-k, -k+1, \ldots, k-1, k\} \text{ for all } i \in [d]\}.$$

(Intuitively, $\mathbf{Q}(\mathbf{w}, k)$ is produced by taking the d-dimensional 'integer box' $[-k, k]^d \cap \mathbb{Z}^d$ and projecting it down to $\mathbb{R}$ via the map $(a_1, \ldots, a_d) \mapsto a_1 w_1 + \cdots + a_d w_d$. This is a specific example of a more general construction called a *generalized arithmetic progression*, and the construction described in Example 1.4 works with $\mathbf{Q}(\mathbf{w}, k)$ replaced with any fixed-dimensional generalized arithmetic progression; we describe it only using $\mathbf{Q}(\mathbf{w}, k)$ for concreteness and to more closely correspond to the statement of Theorem 1.5.)

Example 1.4. Suppose that $v_1, \ldots, v_n$ are all contained in $\mathbf{Q}(\mathbf{w}, k)$, for some $\mathbf{w} = (w_1, \ldots, w_d)$ and $k \in \mathbb{N}$ (think of d as a constant and k as a small power of n). Then for $\varepsilon_1, \ldots, \varepsilon_n \in \{-1, 1\}$, we always have $\sum \varepsilon_i v_i \in \mathbf{Q}(\mathbf{w}, nk)$, which means

$$p(\mathbf{v}) \geq \frac{1}{|\mathbf{Q}(\mathbf{w}, nk)|} \geq \frac{1}{(2nk+1)^d}.$$

Motivated by this example, we would like to say that if $p(\mathbf{v})$ is large, then most of $v_1, \ldots, v_n$ lie in a 'small' generalized arithmetic progression (i.e., one of fixed dimension and $1/\mathsf{poly}(n)$ volume). (We need *most* rather than *all* in such a statement because one can add a small number of arbitrary terms to $\mathbf{v}$ without affecting $p(\mathbf{v})$ too much — specifically, adding $O(\log n)$ terms can only change $p(\mathbf{v})$ by a polynomial factor.)

Tao and Vu [8] prove several statements along these lines. In this paper, we explain the proof of their first inverse theorem.

Theorem 1.5 (Tao–Vu 2009)**.** *For every $d \in \mathbb{N}$, there is a constant $C > 0$ (only depending on d) such that for any $\mathbf{v} = (v_1, \ldots, v_n)$ and $k \in \mathbb{N}$ such that $p(\mathbf{v}) \geq Ck^{-d}$, there exists $\mathbf{w} = (w_1, \ldots, w_r) \in \mathbb{R}^r$ such that $r \leq d - 1$ and all entries of $\mathbf{w}$ are also contained in $\mathbf{v}$, and $v_i \in \bigcup_{a \in [k]} \frac{1}{a}\mathbf{Q}(\mathbf{w}, k)$ for all but at most k^2 indices $i \in [n]$.*

We think of d as fixed and k as a small power of n; then this theorem states that if $p(\mathbf{v}) \geq 1/\mathsf{poly}(n)$, then we can find a projected integer box $\mathbf{Q}(\mathbf{w}, k)$ of constant dimension such that nearly all entries of $\mathbf{v}$ are contained in one of k dilates of $\mathbf{Q}(\mathbf{w}, k)$.

We prove Theorem 1.2 in Section 2 (based on the exposition in [4, Sections 10–12]) and Theorem 1.5 in Section 3 (based on [8, Sections 5–6] from the original paper).

2 Proof of Theorem 1.2

In this section, we prove Theorem 1.2 (of Halász). We first give an outline of the proof in Subsection 2.1, where we describe the main steps of the proof — stated as lemmas — and explain how these lemmas imply Theorem 1.2; in the remaining subsections, we prove these lemmas one at a time.

2.1 Proof Outline

The proof of Theorem 1.2 uses Fourier analytic methods. For this, it'll be convenient to assume that $v_1, \dots, v_n$ are integers. We can make this assumption without loss of generality — given arbitrary real numbers $v_1, \dots, v_n$, we can write down all the equations of the form $\xi_1 v_1 + \cdots + \xi_n v_n = 0$ with $\xi_i \in \{-2, -1, 0, 1, 2\}$ that $\mathbf{v} = (v_1, \dots, v_n)$ satisfy, producing a (massive) system of equations. Then we can obtain a generic solution $\mathbf{v}' = (v_1', \dots, v_n')$ to this system of equations with $v_1', \dots, v_n' \in \mathbb{Z}$ by solving the system over $\mathbb{Q}$ and rescaling. Then $\mathbf{v}'$ satisfies all the same equations of this form that $\mathbf{v}$ does but no others, so $p(\mathbf{v}') = p(\mathbf{v})$.

Throughout the rest of the proof, we'll assume $v_1, \dots, v_n$ are integers, and we'll treat $\mathbf{v}$ as fixed (i.e., we won't quantify over $\mathbf{v}$ in the statements of the following lemmas).

The first step is to use Fourier analysis to bound $p(\mathbf{v})$ by the integral of a continuous function of $t \in [0, 1]$, as given by the following lemma.

Lemma 2.1. *We have* $p_\mu(\mathbf{v}) \le \int_0^1 \prod_{j=1}^n |\cos(2\pi v_j t)| \, dt$.

This transforms the problem of bounding $p(\mathbf{v})$ — which is quite difficult to deal with — into one of bounding the expression $\prod_{j=1}^n |\cos(2\pi v_j t)|$ for each $t \in [0, 1]$, which is much more tractable. However, this expression is somewhat unwieldy, so in the next step, we approximate it by a nicer function and 'group together' values of t based on their contribution to the integral in Lemma 2.1. Specifically, for any $x \in \mathbb{R}$, we define $\|x\| = \min_{m \in \mathbb{Z}} |x - m|$ as the distance from x to the closest integer. Then for each $s \in (0, \infty)$ we define

$$A_s = \{t \in [0, 1] \mid \|v_1 t\|^2 + \cdots + \|v_n t\|^2 \le s\}.$$

We then get the following bound on the right-hand side of Lemma 2.1 (where for a set $A \subseteq [0, 1]$, we use $\lambda(A)$ to denote its Lebesgue measure).

Lemma 2.2. *We have* $\int_0^1 \prod_{j=1}^n |\cos(2\pi v_j t)| \, dt \le \int_0^\infty \lambda(A_s) e^{-s} \, ds$.

Now it remains to bound the measures $\lambda(A_s)$. Intuitively, we only care about the case where s is 'small,' because the factor of e^{-s} means that 'large' s have very small contribution to the integral on the right-hand side of Lemma 2.2 (if s is linear in n, then its contribution to the integral is exponentially small — much smaller than the bound of $n^{-2r-1/2}$ that we are trying to prove). However, it turns out that it is difficult to *directly* bound $\lambda(A_s)$ when s is small. Instead, we use the following lemma to bound $\lambda(A_s)$ for small s in terms of $\lambda(A_s)$ for larger s.

Lemma 2.3. *For all* $s \in (0, \infty)$ *and* $m \in \mathbb{N}$, *we have* $\lambda(A_{m^2 s}) \ge \min\{m\lambda(A_s), 1\}$.

This means it suffices to consider the case where s is reasonably large (specifically, we'll take s to be a small constant times n). In this case, we prove the following bound — this is the step of the argument in which we use the fact that $\mathbf{v}$ satisfies only R linear relations of the specified form.

Lemma 2.4. *We have* $\lambda(A_{n/64}) \le Rn^{-2r}$.

Together, these lemmas immediately imply Theorem 1.2.

Proof of Theorem 1.2. First, we can assume $Rn^{-2r} < 1$ — otherwise the desired statement is immediate from Erdős's bound (1). Then combining Lemmas 2.1 and 2.2, we have

$$p_\mu(\mathbf{v}) \le \int_0^\infty \lambda(A_s) e^{-s} \, ds.$$

First, values of s with $s \geq \frac{n}{64}$ have very little contribution to this integral — we have

$$\int_{n/64}^{\infty} \lambda(A_s) e^{-s}\, ds \leq \int_{n/64}^{\infty} e^{-s}\, ds = e^{-n/64} = O_r(n^{-2r-1/2}) \tag{3}$$

(using the crude bound $\lambda(A_s) \leq 1$). Meanwhile, for each $0 < s \leq \frac{n}{64}$, we'll use Lemma 2.3 to bound $\lambda(A_s)$ in terms of $\lambda(A_{n/64})$ — given s, we can find $m \in \mathbb{N}$ such that $\frac{n}{64} \leq m^2 s \leq \frac{n}{16}$. Then by Lemma 2.4 we have

$$\lambda(A_{m^2 s}) \leq \lambda(A_{n/64}) \leq R n^{-2r}.$$

In particular, since we assumed $Rn^{-2r} < 1$, Lemma 2.3 gives

$$\lambda(A_s) \leq \frac{1}{m}\lambda(A_{m^2 s}) \leq \frac{4s^{1/2}}{n^{1/2}} \cdot R n^{-2r} = 4R s^{1/2} n^{-2r-1/2}.$$

Finally, integrating over s, we have

$$\int_0^{n/64} \lambda(A_s) e^{-s}\, ds \leq 4R n^{-2r-1/2} \int_0^{n/64} s^{1/2} e^{-s}\, ds = O(R n^{-2r-1/2}) \tag{4}$$

(as $\int_0^{\infty} s^{1/2} e^{-s}\, ds$ is finite). Combining (3) and (4) gives the desired bound. □

In the rest of this section, we'll prove each of these lemmas — we'll prove Lemma 2.1 in Subsection 2.2, Lemma 2.2 in Subsection 2.3, Lemma 2.3 in Subsection 2.4, and Lemma 2.4 in Subsection 2.5.

2.2 Proof of Lemma 2.1 — Fourier analysis

The proof of Lemma 2.1 is by Fourier analysis — specifically, we use the fact that for any integer v, we have

$$\int_0^1 e^{2\pi i v t}\, dt = \begin{cases} 1 & \text{if } v = 0 \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Proof of Lemma 2.1. We assumed $v_1, \ldots, v_n$ are all integers, so $\sum \varepsilon_i v_i$ only takes on integer values; then for every integer a, by (5) we have

$$\mathbb{P}[\varepsilon_1 v_1 + \cdots + \varepsilon_n v_n = a] = \mathbb{E}\left[\int_0^1 e^{2\pi i(\varepsilon_1 v_1 + \cdots + \varepsilon_n v_n - a)t}\, dt\right] = \int_0^1 \mathbb{E}[e^{2\pi i(\varepsilon_1 v_1 + \cdots + \varepsilon_n v_n - a)t}]\, dt.$$

Since $\varepsilon_1, \ldots, \varepsilon_n$ are independent, for every $t \in [0,1]$ we can expand

$$\mathbb{E}[e^{2\pi i(\varepsilon_1 v_1 + \cdots + \varepsilon_n v_n - a)t}] = e^{-2\pi i a t} \prod_{j=1}^{n} \mathbb{E}[e^{2\pi i \varepsilon_j v_j t}] = e^{-2\pi i a t} \prod_{j=1}^{n} \cos(2\pi v_j t).$$

Finally, by the triangle inequality this means

$$\mathbb{P}[\varepsilon_1 v_1 + \cdots + \varepsilon_n v_n = a] \leq \int_0^1 \prod_{j=1}^{t} |\cos(2\pi v_j t)|\, dt$$

(since $|e^{2\pi i a t}|$ is always 1). This is true for all integers a, so $p(\mathbf{v}) \leq \int_0^1 \prod_{j=1}^{t} |\cos(2\pi v_j t)|\, dt$ as well. □

2.3 Proof of Lemma 2.2 — simplifying the integral

In order to prove Lemma 2.2, we need the following bound on the cosine function.

Fact 2.5. For all $x \in \mathbb{R}$, we have $|\cos(2\pi x)| \le \exp(-\|2x\|^2)$.

We won't prove this. But heuristically, the reason it's true is that $|\cos(2\pi x)|$ is close to 1 when $2x$ is close to an integer, and if x is close to 0 then $|\cos(2\pi x)| \approx 1 - \frac{1}{2}(2\pi x)^2 \approx \exp(-\frac{1}{2}(2\pi x)^2)$ — and Fact 2.5 has enough slack in the coefficients of x (4 compared to $2\pi^2$) to turn this approximation into a true inequality.

Proof of Lemma 2.2. First, we can imagine choosing $t \sim \text{UNIF}[0,1]$; then the integral we're trying to bound can be written as $\mathbb{E}_t \prod_{j=1}^n |\cos(2\pi v_j t)|$, and using the fact that $\mathbb{E}[X] = \int_0^\infty \mathbb{P}[X \ge x]\,dx$ for any nonnegative random variable X, we can rewrite it as $\int_0^\infty \mathbb{P}_t[\prod_{j=1}^n |\cos(2\pi v_j t)| \ge u]\,du$. Now by Fact 2.5 we have

$$\prod_{j=0}^n |\cos(2\pi v_j t)| \le \exp(-\|2v_1 t\|^2 - \cdots - \|2v_n t\|^2)$$

for all t, which means $\mathbb{P}_t[\prod_{j=1}^n |\cos(2\pi v_j t)| \ge e^{-s}] \le \mathbb{P}_t[\{2t\} \in A_s] = \lambda(A_s)$ for all s (where $\{2t\}$ denotes the fractional part of $2t$, which is also uniformly distributed in $[0,1]$). Substituting $u = e^{-s}$ and plugging this in gives the desired bound. □

2.4 Proof of Lemma 2.3 — from small s to large s

Next we'll prove Lemma 2.3, a lower bound on $\lambda(A_{m^2 s})$ in terms of $\lambda(A_s)$. For this, we'll view the sets A_s as subsets of $\mathbb{R}/\mathbb{Z}$ rather than $[0,1]$ (this makes sense because for $t \in \mathbb{R}$ and $v \in \mathbb{Z}$, the value of $\|vt\|$ only depends on $\{t\}$). For $A, B \subseteq \mathbb{R}/\mathbb{Z}$, we use $A + B$ to denote their sumset

$$A + B = \{a + b \mid a \in A, b \in B\} \subseteq \mathbb{R}/\mathbb{Z},$$

and we use mA to denote the iterated sumset

$$mA = \underbrace{A + \cdots + A}_{m \text{ copies}}.$$

The key observation that goes into the proof of Lemma 2.3 is the following.

Claim 2.6. We have $mA_s \subseteq A_{m^2 s}$.

Proof. We need to show that for any $t_1, \ldots, t_m$ satisfying $\sum_{j=1}^n \|v_j t_i\|^2 \le s$ for every $i \in [m]$, their sum $t = t_1 + \cdots + t_m$ satisfies $\sum_{j=1}^n \|v_j t\|^2 \le m^2 s$. To see this, note that $\|\cdot\|$ satisfies the triangle inequality, so

$$\|v_j t\|^2 \le (\|v_j t_1\| + \cdots + \|v_j t_m\|)^2 \le m(\|v_j t_1\|^2 + \cdots + \|v_j t_m\|^2)$$

for each $j \in [n]$ (the latter inequality is Cauchy–Schwarz). Summing over all $j \in [n]$ gives

$$\sum_{j=1}^n \|v_j t\|^2 \le m \sum_{i=1}^m \sum_{j=1}^n \|v_j t_i\|^2 \le m^2 s$$

(since $\sum_{j=1}^n \|v_j t_i\|^2 \le s$ for each i, and we sum over m values of i). □

In order to use Claim 2.6 to bound $\lambda(A_{m^2 s})$ in terms of $\lambda(A_s)$, we need a version of the Cauchy–Davenport theorem for subsets of $[0,1]$. First, the usual Cauchy–Davenport theorem for $\mathbb{Z}/q\mathbb{Z}$ (where q is a prime) is as follows (we will not prove it; a proof can be found in [7, Section 5.1]).

Theorem 2.7 (Cauchy–Davenport)**.** *For any prime q and nonempty sets $A, B \subseteq \mathbb{Z}/q\mathbb{Z}$, we have*

$$|A+B| \geq \min\{q, |A|+|B|-1\}.$$

We can straightforwardly derive an analogous statement for $\mathbb{R}/\mathbb{Z}$ from the one for $\mathbb{Z}/p\mathbb{Z}$ in the case where A and B are unions of finitely many intervals; this suffices for our purposes (as A_s is the union of finitely many intervals, and if A and B are unions of finitely many intervals then so is $A+B$).

Claim 2.8. For any nonempty sets $A, B \subseteq \mathbb{R}/\mathbb{Z}$ which are each a union of finitely many intervals, we have

$$\lambda(A+B) \geq \min\{\lambda(A)+\lambda(B), 1\}.$$

Proof. We'll convert A and B to subsets of $\mathbb{Z}/q\mathbb{Z}$ for a large prime q and apply Theorem 2.7 to these sets. For any prime q, we can define

$$A^* = \left\{ a \in \{0, \dots, q-1\} \;\middle|\; \left[\frac{a}{q}, \frac{a+1}{q}\right) \subseteq A \right\},$$

and we can define B^* analogously for B. If A and B are each a union of at most k disjoint intervals, then

$$\frac{1}{q}|A^*| \geq |A| - \frac{2k}{q}$$

(and the analogous statement is true for B) — this is because for any interval $[x, y] \subseteq A$, we must have $[x+\frac{1}{q}, y-\frac{1}{q}) \subseteq \bigcup_{a \in A^*} [\frac{a}{q}, \frac{a+1}{q})$ (so the right-hand side, which has measure $\frac{1}{q}|A^*|$, is missing at most a measure-$\frac{2k}{q}$ portion of each of the k intervals that make up A).

Now by Cauchy–Davenport in $\mathbb{Z}/q\mathbb{Z}$, we have $|A^*+B^*| \geq \min\{|A^*|+|B^*|-1, q\}$. Meanwhile, if $a \in A^*$ and $b \in B^*$ then $[\frac{a+b}{q}, \frac{a+b+1}{q}) \subseteq A+B$, so

$$\lambda(A+B) \geq \frac{1}{q}|A^*+B^*| \geq \min\left\{\lambda(A)+\lambda(B)-\frac{4k+1}{q}, 1\right\}.$$

This is true for every prime q, so taking $q \to \infty$ gives the desired result. □

Combining Claims 2.6 and 2.8 immediately gives Lemma 2.3.

2.5 Proof of Lemma 2.4 — a bound for large s

Finally, it remains to prove Lemma 2.4. For this, it'll be convenient to again imagine choosing $t \sim \text{Unif}[0,1]$; then we want to show that

$$\mathbb{P}_t\left[\|v_1 t\|^2 + \cdots + \|v_n t\|^2 \leq \frac{n}{64}\right] \leq Rn^{-2r}.$$

The idea is that we'll convert this inequality into an inequality on a *sum* of cosines — if $\sum_{j=1}^n \|v_j t\|^2$ is small, then $\sum_{j=1}^n \cos(2\pi v_j t)$ must be large. We'll then use Markov's inequality on the $2r$th moment of this sum to upper-bound the probability this occurs (and this moment calculation is where R comes into the picture).

Claim 2.9. If $\sum_{j=1}^n \|v_j t\|^2 \leq \frac{n}{64}$, then $\sum_{j=1}^n \cos(2\pi v_j t) \geq \frac{n}{2}$.

Proof. For all $x \in \mathbb{R}$, we have $\cos(2\pi x) \geq 1 - 32\|x\|^2$ (this is true for similar reasons as Fact 2.5). Summing this over all j gives

$$\sum_{j=1}^n \cos(2\pi v_j t) \geq n - 32\sum_{j=1}^n \|v_j t\|^2 \geq n - 32 \cdot \frac{n}{64} = \frac{n}{2}.$$

□

In order to use Markov's inequality as described earlier, we need the following $2r$th moment computation.

Claim 2.10. We have $\mathbb{E}_t[(\sum_{j=1}^n \cos(2\pi v_j t))^{2r}] = 2^{-2r}R$ (for $t \sim \text{Unif}[0,1]$).

Proof. If we write out each cosine as $\cos x = \frac{1}{2}(e^{ix} + e^{-ix})$ and expand, we get

$$\left(\sum_{j=1}^n \cos(2\pi v_j t)\right)^{2r} = \sum_{j_1,\dots,j_{2r}\in[n]} \sum_{\xi_1,\dots,\xi_{2r}\in\{\pm 1\}} 2^{-2r} e^{2\pi(\xi_1 v_{j_1}+\cdots+\xi_{2r}v_{j_{2r}})t} \tag{6}$$

(where j_1 and ξ_1 correspond to the term $\frac{1}{2}e^{\pm 2\pi i v_j t}$ we take from the first factor when expanding, j_2 and ξ_2 correspond to the term we take from the second factor, and so on). But as seen earlier, for any (fixed) $v \in \mathbb{Z}$ and $t \sim \text{Unif}[0,1]$, we have

$$\mathbb{E}_t[e^{2\pi v t}] = \begin{cases} 1 & \text{if } v = 0 \\ 0 & \text{otherwise} \end{cases}.$$

So when we take the expectation of (6), all terms with $\xi_1 v_{j_1} + \cdots + \xi_{2r} v_{j_{2r}} = 0$ contribute 2^{-2r}, and all other terms contribute 0. But there are precisely R terms with $\xi_1 v_{j_1} + \cdots + \xi_{2r} v_{j_{2r}} = 0$ (by the definition of R), so this expectation is $2^{-2r}R$. □

With this, we're ready to deduce Lemma 2.4.

Proof of Lemma 2.4. By Claim 2.9, we have

$$\lambda(A_{n/64}) = \mathbb{P}_t\left[\|v_1 t\|^2 + \cdots + \|v_n t\|^2 \le \frac{n}{64}\right] \le \mathbb{P}_t\left[\cos(2\pi v_1 t) + \cdots + \cos(2\pi v_n t) \ge \frac{n}{2}\right]$$

(for $t \sim \text{Unif}[0,1]$). Then by Markov's inequality on $(\sum_{j=1}^n \cos(2\pi v_j t))^{2r}$ (which is always nonnegative),

$$\begin{aligned}\mathbb{P}_t\left[\cos(2\pi v_1 t) + \cdots + \cos(2\pi v_n t) \ge \frac{n}{2}\right] &\le \mathbb{P}_t\left[(\cos(2\pi v_1 t) + \cdots + \cos(2\pi v_n t))^{2r} \ge 2^{-2r}n^{2r}\right] \\ &\le \frac{\mathbb{E}[(\sum_{j=1}^n \cos(2\pi v_j t))^{2r}]}{2^{-2r}n^{2r}} \\ &= Rn^{-2r},\end{aligned}$$

where the last equality is by Claim 2.10; this gives the desired bound. □

So we've now proven all four lemmas that go into Theorem 1.2, wrapping up its proof.

3 Proof of Theorem 1.5

In this section, we'll prove Theorem 1.5 (of Tao and Vu). For this proof, it'll be convenient to work in a slightly more general setup, where instead of choosing $\varepsilon_1, \dots, \varepsilon_n \in \{-1, 1\}$, we allow them to be 0 with a certain probability as well.

Definition 3.1. For $\mu \in (0,1]$, we write $\varepsilon \sim \mathcal{P}_\mu$ to denote that ε takes the value 0 with probability $1-\mu$, 1 with probability $\frac{1}{2}\mu$, and -1 with probability $\frac{1}{2}\mu$.

Definition 3.2. For $\mathbf{v} = (v_1, \dots, v_n)$ and $\mu \in (0,1]$, we define $p_\mu(\mathbf{v}) = \max_{a\in\mathbb{R}} \mathbb{P}[\varepsilon_1 v_1 + \cdots + \varepsilon_n v_n = a]$ where $\varepsilon_1, \dots, \varepsilon_n \sim \mathcal{P}_\mu$ are chosen independently.

We're mostly interested in the case $\mu = 1$ (we have $p_1(\mathbf{v}) = p(\mathbf{v})$), but the proof of the theorem will go through smaller values of μ. We'll actually prove the following generalization of Theorem 1.5.

Theorem 3.3 (Tao–Vu 2009)**.** *For every $d \in \mathbb{N}$ and $\mu \in (0, 1]$, there is a constant $C > 0$ (only depending on d and μ) such that for any $\mathbf{v} = (v_1, \dots, v_n)$ and $k \in \mathbb{N}$ such that $p_\mu(\mathbf{v}) \geq Ck^{-d}$, there exists $\mathbf{w} = (w_1, \dots, w_r) \in \mathbb{R}^r$ such that $r \leq d-1$ and all entries of $\mathbf{w}$ are also contained in $\mathbf{v}$, and $v_i \in \bigcup_{a \in [k]} \frac{1}{a}\mathbf{Q}(\mathbf{w}, k)$ for all but at most k^2 indices $i \in [n]$.*

For the proof, it'll be convenient to think of $\mathbf{v}$ as a 'word' — so we'll write $\mathbf{v}$ as $v_1 \dots v_n$ rather than $(v_1, \dots, v_n)$. We'll use $\mathbf{vw}$ to denote the concatenation of two words $\mathbf{v}$ and $\mathbf{w}$, and $\mathbf{v}^m$ to denote the m-fold repetition of $\mathbf{v}$.

We'll also need the following definition.

Definition 3.4. For $\mathbf{w} = (w_1, \dots, w_r) \in \mathbb{R}^r$ and $k \in \mathbb{N}$, we say $\mathbf{w}$ is *k-dissociated* if

$$a_1 w_1 + \dots + a_r w_r \neq 0$$

for all $a_i \in \{-k, -k+1, \dots, k-1, k\}$ with $(a_1, \dots, a_r) \neq (0, \dots, 0)$.

At a high level, the proof of Theorem 3.3 works by algorithmically constructing a k-dissociated word $\mathbf{w}$ from $\mathbf{v}$ one entry at a time; we use certain properties of how $p_\mu(\mathbf{v})$ behaves with respect to word operations (i.e., concatenation and repetition) to ensure that if $p_\mu(\mathbf{v})$ is reasonably large, then so is $p_{\mu'}(\mathbf{w}^{k^2})$ for some μ'. We'll show that if at any step we can't extend $\mathbf{w}$ further, then the current value of $\mathbf{w}$ has the properties we want in Theorem 3.3. Meanwhile, we'll show that if the algorithm runs for long enough that the length of $\mathbf{w}$ becomes d, then $p_{\mu'}(\mathbf{w}^{k^2})$ is small (using the fact that $\mathbf{w}$ is k-dissociated); this will contradict the assumption that $p_\mu(\mathbf{v})$ is reasonably large.

The organization of the rest of the proof is as follows. In Subsection 3.1 we'll state and prove the properties of how $p_\mu(\mathbf{v})$ behaves with respect to word operations that we'll need for the proof. In Subsections 3.2 and 3.3 we'll state and prove the facts that let us deduce that $p_{\mu'}(\mathbf{w}^{k^2})$ is small (if $\mathbf{w}$ is k-dissociated and has length at least d). Finally, in Subsection 3.4 we'll prove Theorem 3.3 by describing the algorithm and showing that it works as in the above sketch.

3.1 Properties of concentration probabilities

We'll first state all the properties of $p_\mu(\mathbf{v})$ that we'll need in one lemma, and then prove them one at a time.

Lemma 3.5. *For all $\mathbf{v} = v_1 \dots v_n$ and $\mathbf{w} = w_1 \dots w_m$ (of lengths n and m), the following properties hold.*

(i) For all $\mu \in (0, 1]$, we have $p_\mu(\mathbf{vw}) \leq p_\mu(\mathbf{v})$.

(ii) For all $\mu \in (0, 1]$, we have $p_\mu(\mathbf{v}) \leq p_{\mu/4}(\mathbf{v})$.

(iii) For all $\mu \in (0, \frac{1}{2}]$ and $d \in \mathbb{N}$, we have $p_\mu(\mathbf{v}) \leq p_{\mu/d}(\mathbf{v}^d)$.

(iv) For all $\mu \in (0, \frac{1}{2}]$, we have $p_\mu(\mathbf{vw}) \leq (\prod_{i=1}^m p_\mu(\mathbf{v}w_i^m))^{1/m}$.

(The purpose of (ii) is simply to allow us to make μ small enough that (iii) and (iv) apply.)

First, (i) can be proven directly from the definition.

Proof of Lemma 3.5(i). By definition, we have $p_\mu(\mathbf{vw}) = \max_{c \in \mathbb{R}} \mathbb{P}[\sum \varepsilon_i v_i + \sum \eta_i w_i = c]$, where $\varepsilon_i, \eta_i \sim \mathcal{P}_\mu$ are all independent; and we can split this as $\max_{c \in \mathbb{R}} (\sum_{a \in \mathbb{R}} \mathbb{P}[\sum \varepsilon_i v_i = a] \mathbb{P}[\sum \eta_i w_i = c - a])$. But for any $c \in \mathbb{R}$, we have $\sum_{a \in \mathbb{R}} \mathbb{P}[\sum \varepsilon_i v_i = a] \mathbb{P}[\sum \eta_i w_i = c - a] \leq \sum_{a \in \mathbb{R}} p_\mu(\mathbf{v}) \mathbb{P}[\sum \eta_i w_i = c - a] = p_\mu(\mathbf{v})$ (because $\mathbb{P}[\sum \varepsilon_i v_i = a] \leq p_\mu(\mathbf{v})$ for all $a \in \mathbb{R}$), as desired. □

In order to prove the remaining properties, we'll use Fourier analysis to obtain an explicit formula for $p_\mu(\mathbf{v})$, similar to Lemma 2.1. First, we can assume $v_1, \ldots, v_n, w_1, \ldots, w_m$ are all integers (just for the proof of Lemma 3.5) for the same reason as in Section 2 — given any real numbers $v_1, \ldots, v_n, w_1, \ldots, w_m$, we can replace them with integers $v_1', \ldots, v_n', w_1', \ldots, w_m'$ satisfying exactly the same set of linear equations of the form relevant to the concentration probabilities in Lemma 3.5. (We'll assume $v_1, \ldots, v_n, w_1, \ldots, w_m$ are integers throughout the rest of this subsection, but not the subsections that follow.)

Claim 3.6. For all $\mu \in (0, 1]$ and $\mathbf{v} = (v_1, \ldots, v_n) \in \mathbb{Z}^n$, we have

$$p_\mu(\mathbf{v}) = \max_{a \in \mathbb{Z}} \int_0^1 e^{-2\pi i a t} \prod_{j=1}^n (1 - \mu + \mu \cos(2\pi v_j t))\, dt.$$

Furthermore, if $\mu \in (0, \frac{1}{2}]$, then this maximum is attained at $a = 0$.

Proof. The proof of the first statement is essentially the same as that of Lemma 2.1, except that now for each $j \in [n]$, for all t we have

$$\mathbb{E}[e^{2\pi i \varepsilon_j v_j t}] = (1 - \mu) \cdot 1 + \frac{1}{2}\mu \cdot e^{2\pi i v_j t} + \frac{1}{2}\mu \cdot e^{-2\pi i v_j t} = 1 - \mu + \mu \cos(2\pi v_j t).$$

For the second statement, note that if $\mu \in (0, \frac{1}{2}]$, then each term $1 - \mu + \mu \cos(2\pi v_j t)$ in the product is nonnegative (for all t), and $|e^{-2\pi i a t}|$ is always 1, so the integral is maximized when $a = 0$ (in which case $e^{-2\pi i a t}$ is 1 for all t). □

We can now obtain the remaining properties in Lemma 3.5 by using this Fourier analytic formula for $p_\mu(\mathbf{v})$ in combination with various inequalities.

Proof of Lemma 3.5(ii). We claim that for all $\mu \in (0, 1]$ and all x, we have

$$|1 - \mu(1 - \cos(x))| \le 1 - \frac{\mu}{4}(1 - \cos(2x)). \tag{7}$$

To see this, note that

$$1 - \cos(2x) = 2\sin^2 x = 2(1 - \cos(x))(1 + \cos(x)) \le 4(1 - \cos(x)),$$

which immediately gives $1 - \mu(1 - \cos(x)) \le 1 - \frac{1}{4}\mu(1 - \cos(2x))$, and $1 - \cos(2x) \le 4(1 + \cos(x))$ for the same reason, which gives $\mu(1 - \cos(x)) - 1 \le 1 - \frac{1}{4}\mu(1 - \cos(2x))$.

Then by using the triangle inequality on our formula for $p_\mu(\mathbf{v})$ from Claim 3.6 and plugging in (7), we get

$$p_\mu(\mathbf{v}) \le \int_0^1 \prod_{j=1}^n |1 - \mu + \mu \cos(2\pi v_j t)|\, dt \le \int_0^1 \prod_{j=1}^n \left(1 - \frac{\mu}{4} + \frac{\mu}{4}\cos(2\pi v_j \cdot 2t)\right) dt = p_{\mu/4}(\mathbf{v}).$$

(We can replace $2t$ with t in the latter integral because the integrand only depends on $\{t\}$.) □

Proof of Lemma 3.5(iii). Using the inequality $1 - dx \le (1 - x)^d$ for $x \in [0, 1]$, we have

$$p_\mu(\mathbf{v}) = \int_0^1 \prod_{j=1}^n (1 - \mu + \mu \cos(2\pi v_j t))\, dt \le \int_0^1 \prod_{j=1}^n \left(1 - \frac{\mu}{d} + \frac{\mu}{d}\cos(2\pi v_j t)\right)^d dt = p_{\mu/d}(\mathbf{v}^d).$$ □

Proof of Lemma 3.5(iii). We use Hölder's inequality (which gives that for any nonnegative functions $f_1, \ldots, f_m$ we have $\int_0^1 f_1 \ldots f_m \le \prod_{i=1}^m (\int_0^1 f_i^m)^{1/m}$) on the functions

$$f_i(t) = \prod_{j=1}^n (1 - \mu + \mu \cos(2\pi v_j t))^{1/m} \cdot (1 - \mu + \mu \cos(2\pi w_i t)).$$

Then we have $\int_0^1 f_1 \ldots f_m = p_\mu(\mathbf{vw})$ and $\int_0^1 f_i^m = p_\mu(\mathbf{v}w_i^m)$ for each $i \in [m]$ (by the Fourier analytic formula in Claim 3.6), so Hölder's inequality gives that $p_\mu(\mathbf{vw}) \le (\prod_{i=1}^m p_\mu(\mathbf{v}w_i^m))^{1/m}$. □

3.2 Anticoncentration for k-dissociated words

In this subsection, we'll prove the following bound on $p_\mu(\mathbf{w}^{k^2})$ when $\mathbf{w}$ is k-dissociated.

Lemma 3.7. *For each $\mu \in (0,1]$, if $\mathbf{w} = w_1 \dots w_d$ is k-dissociated, then $p_\mu(\mathbf{w}) = O_{\mu,d}(k^{-d})$.*

In order to prove this, we can first write

$$p_\mu(\mathbf{w}^{k^2}) = \max_{a\in\mathbb{R}} \mathbb{P}\left[\sum_{i=1}^{d}\sum_{j=1}^{k^2} \varepsilon_{ij} w_i = a\right]$$

for independent $\varepsilon_{ij} \sim \mathcal{P}_\mu$. So for $m \in \mathbb{N}$, we define $\mathcal{Q}_{\mu,m}$ to be the distribution of $\sum_{j=1}^{m} \varepsilon_j$ for independent $\varepsilon_j \sim \mathcal{P}_\mu$ (note that this is always an integer); then we have

$$p_\mu(\mathbf{w}^{k^2}) = \max_{a\in\mathbb{R}} \mathbb{P}[Y_1 w_1 + \dots + Y_d w_d = a]$$

for independent $Y_i \sim \mathcal{Q}_{\mu,k^2}$. In order to prove Lemma 3.7, we'll need the following fact about $\mathcal{Q}_{\mu,m}$.

Lemma 3.8. *For $\mu \in (0, \frac{1}{2}]$ and $m \in \mathbb{N}$, let $Y \sim \mathcal{Q}_{\mu,m}$. Then letting $\sigma = (\mu m)^{1/2}$, for all $a \in \mathbb{Z}$, we have*

$$\mathbb{P}[Y = a] \le 4\sigma^{-1} \cdot \mathbb{P}[Y \in [a - 2\sigma, a + 2\sigma]].$$

Very loosely speaking, this lemma states that $\mathcal{Q}_{\mu,m}$ is roughly 'evenly spread out' on intervals of length roughly σ — the interval $[a-2\sigma, a+2\sigma]$ contains around 4σ integers, so the *average* value of $\mathbb{P}[Y = b]$ over all b in this interval is $\frac{1}{4}\sigma^{-1}\mathbb{P}[Y \in [a-2\sigma, a+2\sigma]]$. Then Lemma 3.8 states that $\mathbb{P}[Y = a]$ cannot be too much (i.e., more than a constant factor) greater than this average.

We defer the proof of this lemma to Subsection 3.3; for now, we'll prove Lemma 3.7 assuming it.

Proof of Lemma 3.7. First, we can assume that $\mu \le \frac{1}{16}$ — otherwise we can replace μ with $\frac{1}{16}\mu$ using Lemma 3.5(ii) twice. We wish to show that $\mathbb{P}[Y_1 w_1 + \dots + Y_d w_d = a] = O_{\mu,d}(k^{-d})$ for all $a \in \mathbb{R}$, where $Y_i \sim \mathcal{Q}_{\mu,k^2}$ are independent. Fix a, and let

$$\mathcal{S} = \{(y_1, \dots, y_d) \in \mathbb{Z}^d \mid y_1 w_1 + \dots + y_d w_d = a\}$$

be the set of 'good' outcomes for $(Y_1, \dots, Y_d)$, so that

$$\mathbb{P}[Y_1 w_1 + \dots + Y_d w_d = a] = \sum_{\mathbf{y}\in\mathcal{S}} \mathbb{P}[Y_i = y_i \text{ for all } i]$$

(where we use $\mathbf{y}$ to denote $(y_1, \dots, y_d) \in \mathbb{Z}^d$). Now we'll use Lemma 3.8 to replace each y_i with a length-k interval around y_i — by Lemma 3.8 we have

$$\mathbb{P}[Y_i = y_i] \le 4\mu^{-1/2}k^{-1} \cdot \mathbb{P}[Y_i \in [y_i - 2\mu^{1/2}k, y_i + 2\mu^{1/2}k]] \le 4\mu^{-1/2}k^{-1} \cdot \mathbb{P}\left[Y_i \in \left[y_i - \frac{1}{2}k, y_i + \frac{1}{2}k\right]\right]$$

for all i and all $y_i \in \mathbb{Z}$ (since $\mu^{1/2} \le \frac{1}{4}$ by assumption). Since $Y_1, \dots, Y_d$ are independent, this means

$$\mathbb{P}[Y_1 w_1 + \dots + Y_d w_d = a] \le 4^d \mu^{-d/2} k^{-d} \sum_{\mathbf{y}\in\mathcal{S}} \mathbb{P}\left[Y_i \in \left[y_i - \frac{1}{2}k, y_i + \frac{1}{2}k\right] \text{ for all } i\right] \tag{8}$$

(we can split $\mathbb{P}[Y_i = y_i \text{ for all } i] = \prod_{i=1}^{d} \mathbb{P}[Y_i = y_i]$, and expanding the points y_i into intervals lets us pick up a factor of $4\mu^{-1/2}k$ for each i).

But using the fact that $\mathbf{w}$ is k-dissociated, we can show that the events that $Y_i \in [y_i - \frac{1}{2}k, y_i + \frac{1}{2}k]$ for all i for different choices of $\mathbf{y} \in \mathcal{S}$ are disjoint — if there were distinct $\mathbf{y}, \mathbf{y}' \in \mathcal{S}$ and integers $b_i, b_i' \in [-\frac{1}{2}k, \frac{1}{2}k]$ such that $y_i + b_i = y_i' + b_i'$ for all i, then since $\sum_i y_i w_i = \sum_i y_i' w_i = a$ (by the definition of $\mathcal{S}$) we would have $\sum_i (b_i - b_i')w_i = 0$, and since $|b_i - b_i'| \le |b_i| + |b_i'| \le k$ for all i, this contradicts the k-dissociativity of $\mathbf{w}$.

So the sum in the right-hand side of (8) is a sum of probabilities of disjoint events, which means it is at most 1; this gives

$$\mathbb{P}[Y_1 w_1 + \dots + Y_d w_d = a] \le 4^d \mu^{-d/2} k^{-d} = O_{\mu,d}(k^{-d}). \qquad \square$$

3.3 Proof of Lemma 3.8

In order to prove Lemma 3.8, we'll first prove a few intermediate results about $Y \sim \mathcal{Q}_{\mu,m}$. We'll assume that $\mu \le \frac{1}{2}$ throughout this subsection (this is mostly just for convenience, and similar arguments can be made even for $\frac{1}{2} \le \mu \le 1$, possibly with different constants). We'll also use σ to denote $(\mu m)^{1/2}$ throughout this subsection. (The reason for this notation is that $\sigma^2 = \mathrm{Var}[Y]$, as we will compute later.)

The first is an anticoncentration bound similar to (1) for the case $\mathbf{v} = 1^m$ (but for general $\mu \le \frac{1}{2}$).

Claim 3.9. For $Y \sim \mathcal{Q}_{\mu,m}$, we have $\mathbb{P}[Y = a] \le \sigma$ for all $a \in \mathbb{Z}$.

Proof. First, we have $\max_{a\in\mathbb{Z}} \mathbb{P}[Y = a] = p_\mu(1^m)$, and by Claim 3.6, we have the explicit formula

$$p_\mu(1^m) = \int_{-1/2}^{1/2} (1 - \mu + \mu\cos(2\pi t))^m \, dt.$$

(The value of the integrand only depends on $\{t\}$, so it doesn't matter whether we integrate over $[0, 1]$ or $[-\frac{1}{2}, \frac{1}{2}]$, and the latter will be slightly more convenient here.) We can now use a similar approximation for the cosine function as in Fact 2.5 — for all $t \in [-\frac{1}{2}, \frac{1}{2}]$ we have $1 - \cos(2\pi t) \ge 4t^2$, which means

$$1 - \mu + \mu\cos(2\pi t) \le 1 - 4\mu t^2 \le e^{-4\mu t^2},$$

and therefore

$$p_\mu(1^m) \le \int_{-1/2}^{1/2} e^{-4\mu m t^2} \, dt \le \int_{-\infty}^{\infty} e^{-4\mu m t^2} \, dt = \frac{1}{2\sigma} \int_{-\infty}^{\infty} e^{-u^2} \, du \le \sigma^{-1}$$

(where we substitute $u = 2\sigma t$ and use the fact that $\int_{-\infty}^{\infty} e^{-u^2} = \sqrt{\pi} \le 2$). □

On the other hand, using Chebyshev's inequality we can show that $Y \sim \mathcal{Q}_{\mu,m}$ is fairly concentrated on an interval of length $O(\sigma)$.

Claim 3.10. For $Y \sim \mathcal{Q}_{\mu,m}$, we have $\mathbb{P}[|Y| \le 2\sigma] \ge \frac{3}{4}$.

Proof. For each $\varepsilon_i \sim \mathcal{P}_\mu$ we have $\mathrm{Var}[\varepsilon_i] = \mathbb{E}[\varepsilon_i^2] = \mu$, which means $\mathrm{Var}[Y] = \mu m = \sigma^2$ (since Y is a sum of m independent variables $\varepsilon_i \sim \mathcal{P}_\mu$). By Chebyshev's inequality, this means $\mathbb{P}[|Y| \ge 2\sigma] \le \frac{1}{4}$. □

The final observation about $Y \sim \mathcal{Q}_{\mu,m}$ we need is that if we fix the parity of a, then $\mathbb{P}[Y = a]$ decreases as a moves further away from 0.

Claim 3.11. For $Y \sim \mathcal{Q}_{\mu,m}$, for all $a \ge 0$, we have $\mathbb{P}[Y = a] \ge \mathbb{P}[Y = a + 2]$.

Proof. Let $Y = \varepsilon_1 + \cdots + \varepsilon_m$ for independent $\varepsilon_i \sim \mathcal{P}_\mu$, and let $S = \{i \in [m] \mid \varepsilon_i \ne 0\}$; we'll show that in fact $\mathbb{P}[Y = a \mid S] \ge \mathbb{P}[Y = a+2 \mid S]$ for all S. Once we condition on S, the variables ε_i for $i \in S$ are independent and uniform in $\{-1, 1\}$ (and all others are 0). If $|S| = b$, then in order to have $Y = a$, exactly $\frac{b+a}{2}$ of them must be 1; similarly, in order to have $Y = a$, exactly $\frac{b+a}{2} + 1$ of them must be 1. (In particular, we must have $b \equiv a \pmod 2$, or else both probabilities are 0). But since $\binom{b}{c}$ is decreasing in c for $c \ge \frac{1}{2}b$, the former is more likely than the latter. □

We can now combine these claims to prove Lemma 3.8.

Proof of Lemma 3.8. Since Y is symmetric, we can assume without loss of generality that $a \geq 0$. If $a \geq 2\sigma$, then we can directly use Claim 3.11 — for all integers $b \in [a - 2\sigma, a]$ with the same parity as a, by Claim 3.11 we have $\mathbb{P}[Y = b] \geq \mathbb{P}[Y = a]$, and since there are at least σ such integers b, we get that

$$\mathbb{P}[Y \in [a - 2\sigma, a + 2\sigma]] \geq \sigma \cdot \mathbb{P}[Y = a].$$

Meanwhile, if $0 \leq a \leq 2\sigma$, then the interval $[a - 2\sigma, a + 2\sigma]$ contains $[0, 2\sigma]$. So on one hand we have $\mathbb{P}[Y = a] \leq \sigma^{-1}$ by Claim 3.9, and on the other hand we have

$$\mathbb{P}[Y \in [a - 2\sigma, a + 2\sigma]] \geq \mathbb{P}[Y \in [0, 2\sigma]] \geq \frac{3}{16} \geq \frac{1}{4}$$

by Claim 3.10 and the fact that Y is symmetric; combining these bounds gives Lemma 3.8. □

Remark 3.12. It's actually possible to use similar ideas to prove the more general statement that

$$\mathbb{P}[Y = y] \lesssim \max\{\tau^{-1}, \sigma^{-1}\}\mathbb{P}[Y \in [y - \tau, y + \tau]]$$

for all $y \in \mathbb{Z}$ and $\tau \in \mathbb{N}$. But since Lemma 3.8 is enough for our purposes, we only prove it (instead of this more general statement) because the proof is simpler.

3.4 The algorithmic construction

Finally, we're ready to prove Theorem 3.3 by constructing $\mathbf{w}$ algorithmically.

Proof of Theorem 3.3. We'll give an algorithm that constructs $\mathbf{w}$ one entry at a time, maintaining the following two invariants:

- At every step, $\mathbf{w}$ is k-dissociated.
- After r steps, when $\mathbf{w}$ has length r, we have $p_\mu(\mathbf{v}) \leq p_{\mu/4d}(\mathbf{v}^{d-r}\mathbf{w}^{k^2})$.

First, we initialize $\mathbf{w} = \emptyset$ and $r = 0$. To see that this satisfies the second invariant, note that $p_\mu(\mathbf{v}) \leq p_{\mu/4}(\mathbf{v}) \leq p_{\mu/4d}(\mathbf{v}^d)$, where the first inequality is by Lemma 3.5(ii) and the second by Lemma 3.5(iii).

Then while $r < d$, we perform the following:

- If the number of indices $i \in [n]$ such that $\mathbf{w}v_i$ is k-dissociated is less than k^2, then terminate and return $\mathbf{w}$. Note that if $\mathbf{w}v_i$ is not k-dissociated, then $v_i \in \frac{1}{a}\mathbf{Q}(\mathbf{w}, k)$ for some $a \in [k]$ — this is because there must exist $a_1, \dots, a_r, a \in \{-k, \dots, k\}$, not all zero, such that

  $$a_1 w_1 + \cdots + a_r w_r + a v_i = 0,$$

  and we must have $a \neq 0$ (otherwise $\{w_1, \dots, w_r\}$ would themselves not be k-dissociated), so

  $$v_i = -\frac{1}{a}(a_1 w_1 + \cdots + a_r w_r) \in \frac{1}{|a|}\mathbf{Q}(\mathbf{w}, k).$$

  So if we reach this step, this means $v_i \in \bigcup_{a \in [k]} \frac{1}{a}\mathbf{Q}(\mathbf{w}, k)$ for all but fewer than k^2 indices i, so $\mathbf{w}$ satisfies the properties in Theorem 3.3, and we're done.
- Otherwise, let $\mathcal{I} = \{i_1, \dots, i_{k^2}\}$ be a set of k^2 indices $i \in [n]$ such that $\mathbf{w}v_i$ is k-dissociated for each $i \in \mathcal{I}$. Our goal is to append v_i to $\mathbf{w}$ for some $i \in \mathcal{I}$. Any choice of $i \in \mathcal{I}$ will satisfy the first invariant, so we wish to find one that also satisfies the second. To do so, note that

  $$p_{\mu/4d}(\mathbf{v}^{d-r}\mathbf{w}^{k^2}) = p_{\mu/4d}(\mathbf{v}^{d-r-1}\mathbf{w}^{k^2} v_1 \dots v_n) \leq p_{\mu/4d}(\mathbf{v}^{d-r-1}\mathbf{w}^{k^2} v_{i_1} \dots v_{i_{k^2}})$$

by Lemma 3.5(i) (we've simply expanded out one of the $d-r$ copies of $\mathbf{v}$ and dropped all the entries that aren't relevant to us), and

$$p_{\mu/4d}(\mathbf{v}^{d-r-1}\mathbf{w}^{k^2}v_{i_1}\dots v_{i_{k^2}}) \le \left(\prod_{i\in\mathcal{I}} p_{\mu/4d}(\mathbf{v}^{d-r-1}\mathbf{w}^{k^2}v_i^{k^2})\right)^{1/k^2}$$

by Lemma 3.5(iv). Combining these gives that there must exist some $i \in \mathcal{I}$ such that

$$p_{\mu/4d}(\mathbf{v}^{d-r}\mathbf{w}^{k^2}) \le p_{\mu/4d}(\mathbf{v}^{d-r-1}\mathbf{w}^{k^2}v_i^{k^2})$$

(since the left-hand side is at most the geometric mean of the right-hand side over all $i \in \mathcal{I}$).

We then append v_i to $\mathbf{w}$ (for this choice of i) — so we replace $\mathbf{w}$ with $\mathbf{w}v_i$ and increment r by 1. This maintains both invariants.

In order to prove Theorem 3.3, it is enough to prove that this algorithm must terminate before r becomes d. Assume not, so that the algorithm produces $\mathbf{w} = w_1 \dots w_d$ of length d satisfying the two invariants. Then by the second invariant, we have $p_{\mu/4d}(\mathbf{w}^{k^2}) \ge p_\mu(\mathbf{v}) \ge Ck^{-d}$ (where C is the constant in the statement of Theorem 3.3, which we'll choose soon). But since $\mathbf{w}$ is k-dissociated, by Lemma 3.7 we must have $p_{\mu/4d}(\mathbf{w}^{k^2}) = O_{\mu,d}(k^{-d})$. If we take C to be a constant (only depending on μ and d) larger than the implicit constant in this bound, then this is a contradiction. □

References

[1] P. Erdős. Extremal problems in number theory. In *Proc. Sympos. Pure Math.*, volume VIII, pages 181–189, Providence, RI, 1965. Amer. Math. Soc.

[2] P. Erdős. On a lemma of Littlewood and Offord. *Bull. Amer. Math. Soc.*, 51(12):898–902, 1945.

[3] G. Halász. Estimates for the concentration function of combinatorial number theory and probability. *Period. Math. Hungar.*, 8(3–4):198–211, 1977.

[4] Manjunath Krishnapur. Anti-concentration inequalities. Lecture notes, https://math.iisc.ac.in/~manju/anti-concentration.pdf.

[5] 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.*, 54(3):277–286, 1943.

[6] A. Sárkőzy and E. Szemerédi. Über ein problem von Erdős und Moser. *Acta Arithmetica*, 11:205–208, 1965.

[7] Terence Tao and Van H. Vu. *Additive combinatorics*. Cambridge University Press, 2006.

[8] Terence Tao and Van H. Vu. Inverse Littlewood–Oxford theorems and the condition number of random discrete matrices. *Annals of Mathematics*, 169:595–632, 2009.