A NEW APPROACH TO AN OLD PROBLEM OF ERDŐS AND MOSER

HOI H. NGUYEN

ABSTRACT. Let $\eta_i, i = 1, \ldots, n$ be iid Bernoulli random variables, taking values ± 1 with probability $\frac{1}{2}$. Given a multiset V of n elements $v_1, \ldots, v_n$ of an additive group G, we define the *concentration probability* of V as

$$\rho(V) := \sup_{v \in G} \mathbf{P}(\eta_1 v_1 + \ldots \eta_n v_n = v).$$

An old result of Erdős and Moser asserts that if v_i are distinct real numbers then $\rho(V)$ is $O(n^{-\frac{3}{2}} \log n)$. This bound was then refined by Sárközy and Szemerédi to $O(n^{-\frac{3}{2}})$, which is sharp up to a constant factor. The ultimate result dues to Stanley who used tools from algebraic geometry to give a complete description for sets having optimal concentration probability; the result now becomes classic in algebraic combinatorics.

In this paper, we will prove that the optimal sets from Stanley's work are stable. More importantly, our result gives an almost complete description for sets having large concentration probability.

1. INTRODUCTION

Let $\eta_i, i = 1, \ldots, n$ be iid Bernoulli random variables, taking values ± 1 with probability $\frac{1}{2}$. Given a multiset V of n elements $v_1, \ldots, v_n$ of an additive group G, we define the random walk S with steps in V to be the random variable $S := \sum_{i=1}^n \eta_i v_i$. The *concentration probability* of V is defined to be

$$\rho(V) := \sup_{v \in G} \mathbf{P}(S = v).$$

Motivated by their study of random polynomials, in the 1940s Littlewood and Offord [5] raised the question of bounding $\rho(V)$ where v_i are real numbers. They showed that $\rho(V) = O(n^{-\frac{1}{2}} \log n)$. Shortly after Littlewood-Offord paper, Erdős [1] gave a beautiful combinatorial proof of the refinement

$$\rho(V) \leq \frac{\binom{n}{n/2}}{2^n} = O(n^{-\frac{1}{2}}). \tag{1}$$

Erdős' result is sharp, as demonstrated by $V = \{1, \ldots, 1\}$.

The Littlewood-Offord and Erdős results are classic in combinatorics and have generated an impressive wave of research, in particular from the early 1960s to the late 1980s.

One direction of research was to generalize Erdős' result to other groups. For example, in 1966 and 1970, Kleitman extended Erdős' result to complex numbers and normed vectors, respectively.

Another direction was motivated by the observation that (1) can be improved significantly under additional assumptions on V. The first such result was discovered by Erdős and Moser [2], who showed that if v_i are distinct real numbers then $\rho(V) = O(n^{-\frac{3}{2}} \log n)$. They conjectured that the logarithmic term is not necessary. Sárközy and Szemerédi claimed this conjecture in the 1960s.

Theorem 1.1 (Sárközy-Szemerédi's theorem, [8])**.** *Let V be a set of n distinct integers, then*

$$\rho(V) = O(n^{-\frac{3}{2}}).$$

Unfortunately, the result of Sárközy and Szemerédi does not cover the most interesting question (also raised by Erdős and Moser [2]): which sets have optimal concentration probability? This major problem remained open until the breakthrough of Stanley [9] in the early 1980s.

Theorem 1.2 (Stanley's theorem)**.** *Let n be odd and $V_0 := \{-\lfloor \frac{n}{2} \rfloor, \ldots, \lfloor \frac{n}{2} \rfloor\}$. Let V be any set of n distinct integers, then*

$$\rho(V) \leq \rho(V_0).$$

A similar result holds for the case n is even [9]. Stanley's proof of Theorem 1.2 used sophisticated machineries from algebraic geometry, in particular the hard-Lepschetz theorem. Few years later, a more elementary proof was given by Proctor [7]. This proof, now becomes classic in algebraic combinatorics, is also of algebraic nature, involving the representation of the Lie algebra $sl(2, \mathbf{C})$.

Although Stanley's result has resolved the question of Erdős and Moser, the algebraic approach has some disadvantage. First, it does not yield the actual value of $\rho(V_0)$. From Theorem 1.1, one would guess (under the assumption that the elements of V are different) that $\rho(V_0) = (C_0 + o(1))n^{-\frac{3}{2}}$ for some constant $C_0 > 0$. The algebraic proofs does not confirm this estimate. (Although one can calculate C_0 by a basic analytical method.)

Assume that C_0 exists for a moment, one would next wonder if V_0 is a stable maximizer. In other words, if some other set V has $\rho(V)$ close to $C_0 n^{-\frac{3}{2}}$,

then should V, in some sense, close to V_0 ? (Notice that ρ is invariant under dilation.)

The goal of this paper is to address these issues.

Theorem 1.3 (Stability theorem). *There is a positive constant ϵ_0 such that for any $0 < \epsilon \le \epsilon_0$, there is a positive number $\epsilon' = \epsilon'(\epsilon)$ such that $\epsilon' \to 0$ as $\epsilon \to 0$, and the following holds: if V is a set of n distinct integers and*

$$\rho(V) \ge (\sqrt{\frac{24}{\pi}} - \epsilon)n^{-\frac{3}{2}},$$

then there is an integer l which divides all $v \in V$ and

$$\sum_{v \in V}(\frac{v}{l})^2 \le (1 + \epsilon')\sum_{v \in V_0}v^2.$$

It follows rather explicitly from the proof of Theorem 1.3 that

$$\rho(V) \le (\sqrt{\frac{24}{\pi}} + o(1))n^{-\frac{3}{2}}, \tag{2}$$

for any set V of n distinct elements of a torsion free group. Here we note that

$$\rho(V_0) = (\sqrt{\frac{24}{\pi}} + o(1))n^{-\frac{3}{2}}. \tag{3}$$

As a byproduct, we obtain the first non-algebraic proof for the asymptotic version of Stanley theorem, and in a more general way: our elements can be from any torsion free groups. More importantly, in case $V \subset \mathbf{Z}$, theorem 1.3 and its proof reveal a natural reason for V_0 to be the optimal set: *This is the set (modulo a dilation) that minimizes the variance $\sum_{v \in V} v^2$ of the random sum S.*

Our method is motivated by a recent idea of Tao and Vu: in certain problems on random discrete matrices we need to *characterize sets of large concentration probability* (under general assumption on V and η_i). The interested reader may also see [6],[12],[13] for more motivations and ideas lying behind this paper.

Theorem 1.3 is proven in two steps. In the first step (Section 3) we give a complete characterization of sets having optimal concentration probability up to a constant factor. In the second step (Section 5) we exploit this strong information to verify the stability part.

2. PRELIMINARY SETTING

The first step is to embed the problem into the finite field $\mathbf{F}_p$ for some prime p. In the case when v_i are integers, we simply take p to be a large prime (for instance $p \geq 2^n(\sum_{i=1}^n |v_i| + 1)$ suffices). If V is a subset of a general torsion-free group G, one can use Theorem B.4 (see Appendix.)

From now on, we can assume that v_i are elements of $\mathbf{F}_p$ for some large prime p. We view elements of $\mathbf{F}_p$ as integers between 0 and $p-1$. We use short hand ρ to denote $\rho(V)$.

Fourier Analysis. The main advantage of working in $\mathbf{F}_p$ is that one can make use of discrete Fourier analysis. Assume that

$$\rho = \rho(V) = \mathbf{P}(S = a),$$

for some $a \in \mathbf{F}_p$. Using the standard notation $e_p(x)$ for $\exp(\frac{2\pi\sqrt{-1}x}{p})$, we have

$$\rho = \mathbf{P}(S = a) = \mathbf{E}\frac{1}{p}\sum_{\xi \in \mathbf{F}_p} e_p(\xi(S-a)) = \mathbf{E}\frac{1}{p}\sum_{\xi \in \mathbf{F}_p} e_p(\xi S)e_p(-\xi a). \qquad (4)$$

By independence

$$\mathbf{E}e_p(\xi S) = \prod_{i=1}^n e_p(\xi\eta_i v_i) = \prod_{i=1}^n \cos\frac{2\pi\xi v_i}{p} \qquad (5)$$

Since $|e_p(-\xi a)| = 1$, it follows that

$$\rho \leq \frac{1}{p}\sum_{\xi \in \mathbf{F}_p} |\cos\frac{2\pi v_i \xi}{p}| = \frac{1}{p}\sum_{\xi \in \mathbf{F}_p} |\frac{\cos \pi v_i \xi}{p}|. \qquad (6)$$

By convexity, we have that $|\sin \pi z| \geq 2\|z\|$ for any $z \in \mathbf{R}$, where $\|z\| := \|z\|_{\mathbf{R}/\mathbf{Z}}$ is the distance of z to the nearest integer. Thus,

$$|\cos\frac{\pi x}{p}| \leq 1 - \frac{1}{2}\sin^2\frac{\pi x}{p} \leq 1 - 2\|\frac{x}{p}\|^2 \leq \exp(-2\|\frac{x}{p}\|^2), \qquad (7)$$

where in the last inequality we used that fact that $1 - y \leq \exp(-y)$ for any $0 \leq y \leq 1$.

Consequently, we obtain a key inequality

$$\rho \le \frac{1}{p} \sum_{\xi \in \mathbf{F}_p} \prod_i |\cos \frac{\pi v_i \xi}{p}| \le \frac{1}{p} \sum_{\xi \in F_p} \exp(-2 \sum_{i=1}^n \|\frac{v_i \xi}{p}\|^2). \qquad (8)$$

Large level sets. Now we consider the level sets $S_m := \{\xi : \sum_{i=1}^n \|\frac{v_i \xi}{p}\|^2 \le m\}$. Decompose $\mathbf{F}_p$ into $\{0\} \cup S_1 \cup (S_2 \backslash S_1) \cup \ldots$, we have

$$n^{-C} \le \rho \le \frac{1}{p} \sum_{\xi \in \mathbf{F}_p} \exp(-2 \sum_{i=1}^n \|\frac{v_i \xi}{p}\|^2) \le \frac{1}{p} + \frac{1}{p} \sum_{m \ge 1} \exp(-2(m-1))|S_m|.$$

Since $\sum_{m \ge 1} \exp(-m) < 1$, there must be is a large level set S_m such that

$$|S_m| \exp(-m+2) \ge \rho p. \qquad (9)$$

We will use (9) as the starting point for our analysis.

For the rest of this section, we show that $S_{\frac{n}{100}}$ can be viewed as some sort of a *dual* set of V. In fact, one can show as far as cardinality is concerned, it does behave like a dual

$$|S_{\frac{n}{100}}| \le \frac{8p}{|V|} = \frac{8p}{n}. \qquad (10)$$

To see this, define $T_a := \sum_{\xi \in V} \cos \frac{2\pi a \xi}{p}$. Using the fact that $\cos 2\pi z \ge 1 - 50\|z\|^2$ for any $z \in \mathbf{R}$, we have, for any $a \in S_{\frac{n}{100}}$

$$T_a \ge \sum_{\xi \in V}(1 - 50\|\frac{a\xi}{p}\|^2) \ge \frac{1}{2}n.$$

One the other hand, using the basic identity

$$\sum_{a \in \mathbf{F}_p} \cos \frac{2\pi a x}{p} = \begin{cases} p & \text{if } x = 0, \\ 0 & \text{otherwise.} \end{cases}$$

We have

$$\sum_{a \in \mathbf{F}_p} T_a^2 \le 2pn.$$

(10) follows from the last two estimates and averaging.

3. A COMPLETE CHARACTERIZATION OF SETS HAVING LARGE CONCENTRATION PROBABILITY

The goal of this section is to demonstrate the fact that as long as $\rho(V) = \Omega(n^{-\frac{3}{2}})$ in $\mathbf{F}_p$, where p is large enough, then most of the elements of V belong to a short interval after an appropriate dilation.

Theorem 3.1 (Characterization of sets having large concentration probability in $\mathbf{F}_p$). *Let δ be a positive constant. Then there are constants $C_1 = C_1(\delta) > 0$ and $C_2 = C_2(\delta) > 0$ such that the following holds. Let $V = \{v_1, \ldots, v_n\}$ be a subset of size n of $\mathbf{F}_p$, where $p \gg n$ is a large prime, such that $\rho(V) \geq \delta n^{-\frac{3}{2}}$. Then there exists a number $k = k(p) \in \mathbf{F}_p$ and a partition $V = V_1 \cup V_2 = k \cdot (W_1 \cup W_2)$ with the following properties:*

- $|W_1| \leq C_1$,
- $\sum_{w \in W_2} w^2 \leq \frac{C_2 n^3}{p^2}$.

The appearance of k is necessary as $\rho(V)$ is invariant under dilation in $\mathbf{F}_p$.

Proof. (of Theorem 3.1) By (9) we have

$$|S_m| \geq \exp(\frac{m}{4} - 2)\rho p \geq \delta \exp(\frac{m}{4} - 2)p n^{-\frac{3}{2}} \qquad (11)$$

for some $m = O(\log n)$ (it will turn out that $m = O(1)$ later on).

Structure of S_m. Consider a set sequence, $S_m, 2S_m, \ldots, 2^l S_m$, where l is the largest integer such that $4^l m \leq \frac{n}{100}$. Assume that i_0 is the smallest index such that $|2^i S_m| \geq 2.1|2^{i-1} S_m|$ for $1 \leq i \leq i_0$ (thus $|2^{i_0} S_m| \geq (2.1)^{i_0}|S_m|$.)

By the definition of level sets, and by Cauchy-Schwarz inequality we observe that $kS_m \subset S_{k^2 m}$ holds for all k. In particular,

$$2^l S_m \subset S_{4^l m} \subset S_{\frac{n}{100}}.$$

On the other hand, by the Cauchy-Davenport theorem (see [14, Theorem 5.4]), we have

$$|2^l S_m| \geq 2^{(l-i_0)}(|2^{i_0} S_m| - 1).$$

Hence,

$$\begin{aligned}
|S_{\frac{n}{100}}| &\geq |2^l S_m| \geq 2^{(l-i_0)}(|2^{i_0} S_m| - 1) \\
&\geq 2^{(l-i_0)}(2.1^{i_0}|S_m| - 1) \\
&\geq 2^{l-1}(\frac{2.1}{2})^{i_0}|S_m| \\
&\geq 2^{l-1}(\frac{2.1}{2})^{i_0}\delta \exp(\frac{m}{4} - 2)pn^{-\frac{3}{2}}.
\end{aligned}$$

Insert the estimate $|S_{\frac{n}{100}}| \leq \frac{8p}{n}$ from (10), and $2^l = \Theta(\sqrt{\frac{n}{m}})$ into the above inequalities, we derive that $m = O_\delta(1)$ and $i_0 = O_\delta(1)$.

Now we consider the set $X := 2^{i_0} S_m$. For $|X| \leq |S_{\frac{n}{100}}| \leq \frac{8p}{n}$, Theorem B.7 implies that X is Freiman-isomorphism of order $2h$ to a subset X' of the integers (h is a sufficiently large constant to be chosen).

Next, since $|2X'| = |2X| \leq 2.1|X| = 2.1|X'|$, we apply Theorem B.6, and then Lemma B.9 (since h is large) to obtain an arithmetic progression P' of rank 1 and of size $|P'| = \Theta(|X'|)$ such that $P' \subset hX'$.

Lifting back to $\mathbf{F}_p$, we conclude that hX contains an arithmetic progression P of rank one and of size $|P| = \Theta(|X|) \gg pn^{-\frac{3}{2}}$ in $\mathbf{F}_p$. Since X is symmetric, we may assume that P is symmetric. Since $hX = h2^{i_0} S_m \subset S_{h^2 4^{i_0} m}$, we conclude that, with $M := h^2 4^{i_0} m = O_\delta(1)$,

$$P \subset hX \subset S_M.$$

We have showed that there exists a constant $C = O_\delta(1)$ and a symmetric arithmetic progression with rank one and with step $k \in \mathbf{F}_p$, $P = \{ik : -Cpn^{-\frac{3}{2}} \leq i \leq Cpn^{-\frac{3}{2}}\}$ such that the following holds for any $\xi \in P$

$$\sum_{i=1}^n \|\frac{\xi v_i}{p}\|^2 \leq M.$$

Set

$$W := k^{-1} \cdot V (\text{ in } \mathbf{F}_p).$$

We then have

$$\sum_{w \in W} \|\frac{jw}{p}\|^2 \leq M$$

for all j from the range $0 \le j \le Cpn^{-\frac{3}{2}}$.

Summing over j we get

$$\sum_{w \in W} \sum_{0 \le j \le Cpn^{-\frac{3}{2}}} \|\frac{jw}{p}\|^2 \le CMpn^{-\frac{3}{2}}. \tag{12}$$

Let C' be a sufficiently large constant, we set

$$W_1 := \{w \in W : \|\frac{w}{p}\| \ge \frac{C'n^{\frac{3}{2}}}{p}\}.$$

Viewing $\mathbf{F}_p$ as $\{0, \ldots, p-1\}$, we observe that if $w \in W_1$ then $C'n^{\frac{3}{2}} \le w \le p - C'n^{\frac{3}{2}}$. Thus, provided that CC' is sufficient large and $0 \le j \le Cpn^{-\frac{3}{2}}$, there are at least $Cpn^{-\frac{3}{2}}/3$ indices j with $jw \in [\frac{p}{4}, \frac{3p}{4}]$ (in $\mathbf{F}_p$). Note that if this is the case, then $\|\frac{jw}{p}\| \ge \frac{1}{4}$. Thus we have

$$\sum_{0 \le j \le Cpn^{-\frac{3}{2}}} \|\frac{jw}{p}\|^2 \ge (\frac{1}{4})^2 Cpn^{-\frac{3}{2}}/3.$$

Summing over $w \in W_1$ we obtain

$$\sum_{w \in W_1} \sum_{0 \le j \le Cpn^{-\frac{3}{2}}} \|\frac{jw}{p}\|^2 \ge \frac{C}{48}|W_1|pn^{-\frac{3}{2}}.$$

Together with (12), this implies that

$$|W_1| \le 48M.$$

Set

$$W_2 := W \backslash W_1.$$

By definition, $\|\frac{jw}{p}\| = j\|\frac{w}{p}\|$ for all $j \le p(2C'n^{\frac{3}{2}})^{-1}$ and for all $w \in W_2$. Thus

$$\sum_{0 \le j \le pn^{-\frac{3}{2}}} \|\frac{jw}{p}\|^2 \ge \sum_{0 \le j \le p(2C'n^{\frac{3}{2}})^{-1}} j^2 \|\frac{w}{p}\|^2$$
$$\ge \frac{1}{64C'^3} \frac{p^3}{n^{9/2}} \|\frac{w}{p}\|^2.$$

Summing over $w \in W_2$ and using (12), we obtain

$$\sum_{w \in W_2} \frac{1}{64C'^3} \frac{p^3}{n^{9/2}} \|\frac{w}{p}\|^2 \le \sum_{w \in W_2} \sum_{0 \le j \le Cpn^{-\frac{3}{2}}} \|\frac{jw}{p}\|^2$$
$$\le CMpn^{-\frac{3}{2}}.$$

Hence,

$$\sum_{w \in W_2} \|\frac{v}{p}\|^2 \le 64CC'^3 M \frac{n^3}{p^2},$$

concluding the proof.

$\square$

Remark 3.2. The idea to divide sumsets into dyadic sequences was first utilized by Szemerédi and Vu in [10].

4. AN ESTIMATE ON $\rho(V_0)$

To give the reader an overview for the proof of Theorem 1.3, we show in this section the estimate (3) for $\rho(V_0)$

$$\rho(V_0) = (\sqrt{\frac{24}{\pi}} + o(1))n^{-\frac{3}{2}}.$$

View the elements of $\mathbf{F}_p$ as integers between $-(p-1)/2$ and $(p-1)/2$, by (4) we have

$$\mathbf{P}(S = 0) = \frac{1}{p} \sum_{\xi \in \mathbf{F}_p} \prod_{i \in V_0} \cos \frac{2\pi i \xi}{p} = \frac{1}{p} \sum_{\xi \in \mathbf{F}_p} \prod_{i \in V_0} \cos \frac{\pi i \xi}{p}. \tag{13}$$

We split this sum into two parts

$$\Sigma_1 := \frac{1}{p} \sum_{\|\frac{\xi}{p}\| \le \frac{\log^2 n}{n^{3/2}}} \prod_{i \in V_0} \cos \frac{\pi i \xi}{p},$$

$$\Sigma_2 := \frac{1}{p} \sum_{\|\frac{\xi}{p}\| > \frac{\log^2 n}{n^{3/2}}} \prod_{i \in V_0} \cos \frac{\pi i \xi}{p}.$$

We are going to show that

Lemma 4.1.

$$\Sigma_1 = \frac{1}{p} \sum_{\|\frac{\xi}{p}\| \le \frac{\log^2 n}{n^{3/2}}} \prod_{i \in V_0} |\cos \frac{\pi i \xi}{p}| = (\sqrt{\frac{24}{\pi}} + o(1))n^{-\frac{3}{2}}.$$

Lemma 4.2.

$$\Sigma_2 \le \frac{1}{p} \sum_{\|\frac{\xi}{p}\| > \frac{\log^2 n}{n^{3/2}}} \prod_{i \in V_0} |\cos \frac{\pi i \xi}{p}| \le n^{-3}.$$

The two lemmas together imply that $\rho(V_0) = (\sqrt{\frac{24}{\pi}} + o(1))n^{-\frac{3}{2}}$.

Proof. (of Lemma 4.1) The first equality is trivial, as all cos are positive in this range of ξ. Viewing ξ as an integer with absolute value at most $n^{-\frac{3}{2}} p \log^2 n$, we have

$$\cos \frac{\pi i \xi}{p} = 1 - (\frac{1}{2} + o(1)) \frac{\pi^2 i^2 \xi^2}{p^2} = \exp\left(-(1/2 + o(1)) \frac{\pi^2 i^2 \xi^2}{p^2}\right).$$

Since $\sum_{i \in V_0} i^2 = (1 + o(1))\frac{n^3}{12}$, it follows that

$$\Sigma_1 = (1 + o(1)) \int_{|x| \le \frac{\log^2 n}{n^{3/2}}} \exp\left(-(1/2 + o(1)) \frac{n^3 \pi^2}{12} x^2\right).$$

Setting $y = \sqrt{\frac{n^3 \pi^2}{12}} x$, changing variables, and using the gaussian identity $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp(-\frac{y^2}{2})dy = 1$, we have

$$\Sigma_1 = (1 + o(1))(\frac{n^3 \pi^2}{12})^{-1/2} \int_{-\infty}^{\infty} \exp(-\frac{y^2}{2})dy = (\sqrt{\frac{24}{\pi}} + o(1))n^{-\frac{3}{2}},$$

completing the proof. □

Proof. (of Lemma 4.2). To prove Lemma 4.2 we use the upper bound from (7). We split the sum into three subsums, according to the magnitude of $\|\frac{\xi}{p}\|$. We make frequently use of the simple fact that if $|i|\|\frac{\xi}{p}\| \leq 1/2$ then $\|\frac{i\xi}{p}\| = |i|\|\frac{\xi}{p}\|$.

Small $\|\frac{\xi}{p}\|$: $\frac{\log^2 n}{n^{3/2}} \leq \|\frac{\xi}{p}\| \leq \frac{1}{n}$. For all i with $|i| \leq \frac{n}{2}$, we have $\|\frac{i\xi}{p}\| = |i|\|\frac{\xi}{p}\|$.

$$\sum_{i \in V_0} \|\frac{i\xi}{p}\|^2 = \sum_{i \in V_0} i^2 \|\frac{\xi}{p}\|^2 \geq \frac{n^3}{12} \frac{\log^4 n}{n^3} \geq 4\log n \tag{14}$$

Thus, the contribution of this subsum is bounded from above by n^{-4}.

Medium $\|\frac{\xi}{p}\|$: $\frac{1}{n} \leq \|\frac{\xi}{p}\| \leq \frac{1}{4}$. By Cauchy-Schwartz, we have

$$\|\frac{i\xi}{p}\|^2 + \|\frac{i'\xi}{p}\|^2 \geq \frac{1}{2}\|\frac{(i-i')\xi}{p}\|^2.$$

For any ξ in this case, let $W(\xi)$ be the set of pairwise disjoint pairs $(i, i') \in V_0$ that maximizes the sum $M(\xi) := \sum_{(i,i') \in W(\xi)} |i - i'|^2$ under the constrain that $|i - i'|\|\frac{\xi}{p}\| \leq \frac{1}{2}$ for all $(i, i') \in W(\xi)$. It is easy to check that $M(\xi) \geq n^c$ for some constant $c > 0$ for all ξ in this case (For more details see Lemma 5.3.) From here one can conclude that the contribution of this subsum is at most $\exp(n^{-\Omega(1)}) = o(n^{-4})$.

Large $\|\frac{\xi}{p}\|$: $\frac{1}{4} \leq \|\frac{\xi}{p}\| \leq \frac{1}{2}$. By Weyl's equidistribution theorem, the number of $i \in V_0$ such that $\|\frac{i\xi}{p}\| \geq \frac{1}{4}$ is approximately $n/2$. Thus, for any ξ in this category

$$\sum_{i \in V_0} \|\frac{i\xi}{p}\|^2 \geq \frac{n}{8}. \tag{15}$$

Hence, the contribution of this subsum is only $\exp(-\Omega(n))$.

$\square$

5. PROOF OF THEOREM 1.3

First, by applying Theorem 3.1, there exists a $k \in \mathbf{F}_p$ and a partition $V = k \cdot (W_1 \cup W_2)$ such that $|W_1| \leq C_1$ and $\sum_{w \in W_2} w^2 \leq C_2 n^3$, where C_1 and C_2 are positive constants depending on ϵ.

Let $C = C(\epsilon)$ be a large positive constant and $c = c(\epsilon)$ be a small positive constant. By setting C, c properly and moving at most ϵn elements of W_2 to W_1, we can assume the following property for W_2.

Property 5.1. W_2 is a *c-irreducible* subset of size at least $(1 - \epsilon)n$ of the interval $[-Cn, Cn]$. In other words, there is no integer $d \geq 2$ which divides all but $c|W_2|$ elements of W_2.

Indeed, after applying Theorem 3.1, because $\sum_{w \in W_2} w^2 \leq C_2 n^3$, all but $\epsilon n/2$ elements of W_2 belong to the interval $[-Cn, Cn]$, provided that C was chosen to be large enough.

To ensure the irreducibility property, we iterate the following process. First, if W_2, viewed as a subset of integers, can be written as $W_2 = d_2 \cdot W_2' \cup X_2'$ with some integer $d_2 \geq 2$ and $|W_2'| \geq |W_2| - c|W_2|$, then we remove these exceptional $c|W_2|$ elements to W_1 and pass to consider W_2', which is a subset of $[-Cn/2, Cn/2]$. Next, if $W_2' = d_2' \cdot W_2'' \cup X_2''$ with some integer $d_2' \geq 2$ and $|W_2''| \geq |W_2'| - c|W_2'|$, then we remove these $c|W_2'|$exceptional elements to W_1 and switch to consider W_2'', which now belongs to $[-Cn/4, Cn/4]$, and keep iterating the process. Observe that whenever reduction occurs, the size of the new set drops by at most cn, meanwhile the size of the intervals they belong to shrinks by a factor of 2. Thus if c was chosen to be smaller than $\epsilon(4(3 + \log_2 C))^{-1}$, then the process must terminate after at most $3 + \log_2 C$ steps, as the set under consideration would then have size at least $n - C_1 - (3 + \log_2 C)cn \geq n - \epsilon n/2$, but the interval containing it would be $[-n/8, n/8]$, a contradiction.

Thus, by applying Theorem 3.1, there exists a $k \in \mathbf{F}_p$ and a partition $V = k \cdot (W_1 \cup W_2)(\bmod p)$ such that

- $|W_1| \leq \epsilon n$,
- W_2 satisfies Property 5.1.

In the next step, we pass to consider the concentration probability of W by the following identity

$$\mathbf{P}(S_V = kv) = \mathbf{P}(S_W = v) = \frac{1}{p} \sum_{\xi \in \mathbf{F}_p} e_p(-\frac{v\xi}{2}) \prod_{w \in W} \cos \frac{\pi w \xi}{p}.$$

We split the sum into two parts,

$$\Sigma_1 := \frac{1}{p} \sum_{\|\frac{\xi}{p}\| \leq \frac{\log^2 n}{n^{3/2}}} e_p(-\frac{v\xi}{2}) \prod_{w \in W} \cos \frac{\pi w \xi}{p}$$

$$\Sigma_2 := \frac{1}{p} \sum_{\|\frac{\xi}{p}\| > \frac{\log^2 n}{n^{3/2}}} e_p(-\frac{v\xi}{2}) \prod_{w \in W} \cos \frac{\pi w \xi}{p}.$$

We are going to exploit the structure of W_2 to show that

Lemma 5.2.

$$\Sigma_2 \leq \frac{1}{p} \sum_{\|\frac{\xi}{p}\| \geq \frac{\log^2 n}{n^{3/2}}} \prod_{w \in W_2} |\cos \frac{\pi w \xi}{p}| \leq n^{-3}.$$

Proof. Making use of (7), we split the sum into three subsums, according to the magnitude of $\|\frac{\xi}{p}\|$.

Small $\|\frac{\xi}{p}\|$: $\frac{\log^2 n}{n^{3/2}} \leq \|\frac{\xi}{p}\| \leq \frac{1}{2Cn}$. Because $W_2 \subset [-Cn, Cn]$, we easily obtain that

$$\sum_{w \in W_2} \|\frac{w\xi}{p}\|^2 = \sum_{w \in W_2} w^2 \|\frac{\xi}{p}\|^2 \geq \frac{(1-\epsilon)^3 n^3}{12} \frac{\log^4 n}{n^3} \geq 4 \log n.$$

Thus the contribution from this part is bounded from above by n^{-4}.

Medium $\|\frac{\xi}{p}\|$: $\frac{1}{2Cn} \leq \|\frac{\xi}{p}\| \leq \frac{1}{64C}$. To handle this part, we first observe the following simple fact.

Lemma 5.3. *Let $a \in \mathbf{F}_p$ be arbitrary. Let $\xi \in \mathbf{F}_p$ and $l > 0$ such that $l\|\frac{\xi}{p}\| \leq \frac{1}{2}$. Then the following holds for any sequence $0 \leq i_1 < \cdots < i_m \leq l$ with $m \geq 4$*

$$\sum_{j=1}^{m} \|\frac{a + i_j \xi}{p}\|^2 \geq \frac{m^3}{48} \|\frac{\xi}{p}\|^2. \tag{16}$$

The proof of Lemma 5.3 is quite simple, we defer it to Appendix A. Now we arrange the elements of W_2 as

$$-Cn < w_1 < w_2 < \cdots < w_{|W_2|} < Cn.$$

Set $l := \frac{1}{2\|\xi/p\|}$. Thus $32C \leq l \leq Cn$. Let i_1 be the largest index such that $w_{i_1} - w_1 \leq l$. We then move on to choose $i_2 > i_1$, the largest index such that $w_{i_2} - w_{i_1+1} \leq l$, and so on. By so, we create several blocks of elements of W_2 with the property that elements of the same block have difference $\leq l$.

Since $w_{i_j+1} - w_{i_{j-1}+1} > l$ for all j, the number of blocks is less than $\frac{2Cn}{l} + 1$. Next, we call a block *short* if it contains no more than $\frac{l}{8C}$ elements of W_2. The total number of elements of W_2 that belong to short blocks is bounded by $(\frac{2Cn}{l} + 1)(\frac{l}{8C}) \leq \frac{|W_2|}{2}$. Hence there are at least $\frac{|W_2|}{2}$ elements that belong to long blocks.

For simplicity, we divide each long block into smaller blocks of exactly $\lfloor \frac{l}{8C} \rfloor$ elements. The number of such uniform blocks is then at least $\frac{1}{2}\frac{|W_2|/2}{l/8C} = \frac{2C|W_2|}{l} \geq \frac{Cn}{l}$. We apply (16) to each block (with $m = \lfloor l/8C \rfloor$), and then sum over the collection of all blocks using the fact that $l = \frac{1}{2\|\xi/p\|}$,

$$\sum_{w \in W_2} \|\frac{w\xi}{p}\|^2 \geq \frac{Cn}{l}\frac{m^3}{48}\|\frac{\xi}{p}\|^2 \gg l^2 n\|\frac{\xi}{p}\|^2 \gg n.$$

Thus the contribution from this part is bounded from above by $\exp(-\Omega(n))$.

Large $\|\frac{\xi}{p}\|$: $\frac{1}{64C} \leq \|\frac{\xi}{p}\| \leq \frac{1}{2}$. Let $\delta := \frac{1}{\log n}$ and

$$W_2' := \{w \in W_2, \|\frac{w\xi}{p}\| \leq \delta\}.$$

Assume for the moment that $|W_2'| \leq (1-c)|W_2|$ for some positive constant $c = c(\epsilon)$. Then

$$\sum_{w \in W_2} \|\frac{w\xi}{p}\|^2 \geq c|W_2|\delta^2 \gg \delta^2 n \gg \frac{n}{\log^2 n},$$

and so the contribution from this part is bounded from above by $\exp(-\Omega(n \log^{-2} n))$.

Thus, it suffices to show that $|W_2'| \leq (1-c)|W_2|$ for some sufficiently small c. Assume otherwise, we will deduce that there exists a nontrivial $d \in \mathbf{Z}$ that divides all the elements of W_2', which contradicts the c-irreducibility assumption of W_2.

To obtain the above contradiction we first use Lemma B.10 to W_2. It is implied that that $kW_2 - kW_2$ contains an arithmetic progression of rank one, $Q = \{id : |i| \leq \gamma n\}$.

Because $Q \subset [-2kCn, 2kCn]$, the step d must be bounded,

$$0 < d \leq \frac{2kC}{\gamma}. \tag{17}$$

Let q be an element of Q. By definition, we can write q as a sum and difference of elements of W_2,

$$q = w_1 + \cdots + w_k - w_{k+1} - \cdots - w_k.$$

Because $\|\frac{w\xi}{p}\| \leq \delta$ for all i, by the triangle inequality we have

$$\|\frac{q\xi}{p}\| \leq 2k\delta.$$

Apply the above estimate for all elements of Q, one obtains

$$\|\frac{id\xi}{p}\| \leq 2k\delta \text{ for all } |i| \leq \gamma n.$$

But as $2k\delta < \frac{1}{4}$, one must have

$$\|\frac{d\xi}{p}\| \leq \frac{2k\delta}{\gamma n}.$$

Next, view $\mathbf{F}_p$ as the interval $[-(p-1)/2, (p-1)/2]$ of $\mathbf{Z}$ and consider ξ as an integer which satisfies $\frac{p}{64C} \leq |\xi| \leq \frac{p}{2}$.

By the above bound of $\|\frac{d\xi}{p}\|$, we can write $d\xi = sp + t$, and so

$$\xi = \frac{sp+t}{d},$$

where s, t are integers with $|t| \leq \frac{2k\delta p}{\gamma n}$.

We now collect some useful information for s and d. Firstly, because ξ has large absolute value, $|\xi| \geq \frac{p}{64C}$, s cannot be zero. Secondly, because $|\xi| \leq p/2$ and t are small, d does not divide s.

Let w' be an arbitrary element of W_2' and consider in $\mathbf{Z}$ the product $w'\xi$,

$$w'\xi = \frac{w'sp}{d} + \frac{w't}{d} = s'p + \frac{t'p}{d} + \frac{w't}{d}, \tag{18}$$

where $w's = s'd + t'$ with $s', t' \in \mathbf{Z}$, and $-\frac{d}{2} \leq t' \leq \frac{d}{2}$.

Because $|w'| \leq Cn$, we have

$$|\frac{w't}{d}| \leq Cn\frac{2k\delta p/\gamma n}{d} = \frac{2Ck\delta}{\gamma}\frac{p}{d} \leq \frac{p}{2d},$$

where in the last inequality we used the fact that δ is small compared to all other quantities.

We next consider two cases according to the value of t'. We will eventually run to contradiction in both cases, thus completing the proof of Lemma 5.2.

Case 1 : $t' \neq 0$. Because $-\frac{d}{2} \leq t' \leq \frac{d}{2}$, we have $\frac{p}{d} \leq |\frac{t'p}{d}| \leq \frac{p}{2}$, and so

$$\frac{p}{2d} \leq |\frac{t'p}{d} + \frac{wt}{d}| \leq \frac{p}{2} + \frac{p}{2d}.$$

Together with (18), this implies that $\|\frac{w'\xi}{p}\| \geq \frac{1}{2d}$. Hence, by the bound of d from (17),

$$\|\frac{w'\xi}{p}\| \geq \frac{\gamma}{4kC} > \delta.$$

This inequality violates the definition of W_2', so Case 1 does not hold as long as δ is sufficiently small.

Case 2: $t' = 0$. It then follows that d divides sw' for all $w' \in W_2'$. But recall that d does not divide s, so all the element of W_2' are divisible by a nontrivial divisor of d. However, this contradicts with the c-irreducibility assumption of W_2.

$\square$

We have shown that the contribution of Σ_2 is negligible. Thus, it suffices to justify Theorem 1.3 from the following assumption

$$\Sigma_1 \geq (1-\epsilon)\sqrt{\frac{24}{\pi}}n^{-\frac{3}{2}} = (1-\epsilon)\sqrt{\frac{12}{\pi^2 n^3}}\int_{-\infty}^{\infty}\exp(-\frac{y^2}{2})dy. \qquad (19)$$

We have

$$\Sigma_1 = \frac{1}{p} \sum_{\|\frac{\xi}{p}\| \leq \frac{\log^2 n}{n^{3/2}}} e_p(-\frac{v\xi}{2}) \prod_{w \in W} \cos \frac{\pi w \xi}{p}$$

$$\leq \frac{1}{p} \sum_{\|\frac{\xi}{p}\| \leq \frac{\log^2 n}{n^{3/2}}} \prod_{w \in W_1} |\cos \frac{\pi w \xi}{p}| \prod_{w \in W_2} \cos \frac{\pi w \xi}{p}$$

$$\leq \frac{1}{p} \sum_{\|\frac{\xi}{p}\| \leq \frac{\log^2 n}{n^{3/2}}} \exp\left(- \sum_{w \in W_1} \|\frac{\xi w}{p}\|^2 \right) f_{W_2}(\xi),$$

where

$$f_{W_2}(\xi) := \prod_{w \in W_2} \cos \frac{\pi w \xi}{p} = \exp\left(-(1/2 + o(1)) \sum_{w \in W_2} \frac{\pi^2 w^2 \xi^2}{p^2} \right).$$

Combining this estimate and the lower bound for Σ_1 we obtain that

$$\sum_{w \in W_2} w^2 \leq (1 + \epsilon) \frac{n^3}{12}. \tag{20}$$

What remains is to show that a similar sum involving the elements of W_1 is negligible. To this end, we first show that the elements of W_1 do not have large absolute values (viewing $\mathbf{F}_p$ as the interval $[-(p-1)/2, (p-1)/2]$ of $\mathbf{Z}$).

Let $C = C(\epsilon)$ be a number so that

$$\int_{|y| \geq C} \exp(-\frac{y^2}{2}) dy = \epsilon. \tag{21}$$

As p and n are large and $|W_2| \geq (1 - \epsilon)n$, we have

$$\frac{1}{p} \sum_{\frac{C}{n^{3/2}} \leq \|\frac{\xi}{p}\| \leq \frac{\log^2 n}{n^{3/2}}} f_{W_2}(\xi) \leq \left(\sum_{w \in W_2} \pi^2 w^2 \right)^{-1/2} \int_{|y| \geq C} \exp(-\frac{y^2}{2}) dy$$

$$< (1 + 4\epsilon)^{-1} \sqrt{\frac{12}{\pi^2 n^3}} \epsilon \leq \epsilon \sqrt{\frac{12}{\pi^2 n^3}} \int_{-\infty}^{\infty} \exp(-\frac{y^2}{2}) dy.$$

It thus follows from (19) that

$$\frac{1}{p}\sum_{\|\frac{\xi}{p}\|\leq\frac{C}{n^{3/2}}}\exp\left(-\sum_{w\in W_1}\|\frac{w\xi}{p}\|^2\right)f_{W_2}(\xi)\geq(1-2\epsilon)\sqrt{\frac{12}{\pi^2n^3}}\int_{-\infty}^{\infty}\exp(-\frac{y^2}{2})dy.$$

(22)

Now we are ready to state our estimate on the elements of W_1.

Lemma 5.4. *View* $\mathbf{F}_p$ *as the interval* $[-(p-1)/2,(p-1)/2]$ *of* $\mathbf{Z}$. *Let* w_0 *be an arbitrary element of* W_1. *Then,*

$$|w_0|\leq\frac{1}{2C}n^{\frac{3}{2}}.$$

Proof. (of Lemma 5.4) Assume for contradiction that $|w_0|>\frac{n^{\frac{3}{2}}}{2C}$. Then the number of $\xi\in[-\frac{Cp}{n^{3/2}},\frac{Cp}{n^{3/2}}]$ satisfying $\frac{1}{8}\leq\|\frac{w_0\xi}{p}\|$ is at least $(3/4-o(1))\frac{2Cp}{n^{3/2}}$. Denote this set by I, then for $\xi\in I$ we have

$$\sum_{w\in W}\|\frac{w\xi}{p}\|^2\geq\|\frac{w_0\xi}{p}\|^2\geq\frac{1}{64}.$$

Notice that the function $f_{W_2}(\xi)$ is decreasing in $|\xi|$ (in the range $-\frac{Cp}{n^{3/2}}\leq\xi\leq\frac{Cp}{n^{3/2}}$) and $|I|\geq Cp/n^{3/2}$, we have

$$\Sigma_1\leq\frac{1}{p}\sum_{\xi\in I}\exp(-\|\frac{w\xi}{p}\|^2)f_{W_2}(\xi)+\frac{1}{p}\sum_{\|\frac{\xi}{p}\|\leq\frac{C}{n^{3/2}},\xi\notin I}\exp(-\|\frac{w\xi}{p}\|^2)f_{W_2}(\xi)$$

$$\leq\frac{1}{p}\sum_{\xi\in I}\exp(-\frac{1}{64})f_{W_2}(\xi)+\frac{1}{p}\sum_{\|\frac{\xi}{p}\|\leq\frac{C}{n^{3/2}},\xi\notin I}f_{W_2}(\xi)$$

$$\leq\frac{1}{p}\sum_{\|\frac{\xi}{p}\|\leq\frac{C}{n^{3/2}}}f_{W_2}(\xi)-(1-\exp(-\frac{1}{64}))\frac{1}{p}\sum_{\frac{C}{2n^{3/2}}\leq\|\frac{\xi}{p}\|\leq\frac{C}{n^{3/2}}}f_{W_2}(\xi).$$

As p and n are large, the above sum is bounded by

$$\Sigma_1\leq(1+4\epsilon)\sqrt{\frac{12}{\pi^2n^3}}\left(\int_{|y|\leq C}\exp(-\frac{y^2}{2})-(1-\exp(-\frac{1}{64}))\int_{C/2\leq|y|\leq C}\exp(-\frac{y^2}{2})dy\right)$$

$$\leq(1+4\epsilon)\sqrt{\frac{12}{\pi^2n^3}}\left(\sqrt{2\pi}-(1-\exp(-\frac{1}{64}))\epsilon^{2/3}\right)$$

where in the last estimate we used the definition (21) of C and the fact that

$$\int_{C/2 \le |y| \le C} \exp(-\frac{y^2}{2})dy \ge \epsilon^{2/3}.$$

By choosing ϵ sufficiently small, the upper bound for σ_1 above is smaller than the lower bound in (19), a contradiction. This concludes the proof of Lemma 5.4. $\qquad\square$

To continue, set

$$\sigma_1 := \sum_{w \in W_1} w^2.$$

By Lemma 5.4, one has $|w| \le \frac{1}{2C}n^{3/2}$ for every $w \in W_1$. Thus if $\|\frac{\xi}{p}\| \le Cn^{-3/2}$, then

$$\sum_{w \in W_1} \|\frac{w\xi}{p}\|^2 = \sum_{w \in W_1} |w|^2\|\frac{\xi}{p}\|^2 = \sigma_1\|\frac{\xi}{p}\|^2.$$

This observation enables us to apply the proof of Lemma 5.4, with w_0^2 replaced by σ_1, to obtain a similar bound for σ_1

$$\sigma_1 = \sum_{w \in W_1} w^2 \le \frac{n^3}{4C^2}. \tag{23}$$

Putting (20) and (23) together, we obtain

$$\sum_{w \in W} w^2 = \sum_{w \in W_1} w^2 + \sum_{w \in W_2} w^2 \le (1 + \epsilon + \frac{3}{C^2})\frac{n^3}{12}$$
$$\le (1 + \epsilon')\frac{n^3}{12}, \tag{24}$$

where $\epsilon' \to 0$ as $\epsilon \to 0$.

Let us stop to record what we have achieved so far. For any sufficiently large p, there exist a nonzero number $k \in \mathbf{F}_p$ and a set W of integers which satisfies (24) and such that $V = k \cdot W(\bmod\ p)$.

On the other hand, note that the number of sets W which satisfy (24) is bounded by a function of n and ϵ'. Hence, there exists a common set $W = \{w_1, \ldots, w_n\}$ for which $v_i = kw_i(\bmod\ p), v_i \in V$ for infinitely many primes p.

As a consequence, for any pair (i, j) one has $p | k(v_i w_j - v_j w_i)$, or equivalently

$$p | v_i w_j - v_j w_i.$$

Because this is true for infinitely many p, one has

$$v_i w_j = v_j w_i (\text{in } \mathbf{Z}).$$

As this identity holds for every pair (i, j), and as W is irreducible (because of either (24) or Property 5.1), there exists a nonzero integer l such that

$$V = l \cdot W.$$

In other words, l divides every element of V and

$$\sum_{v \in V} (\frac{v}{l})^2 = \sum_{w \in W} w^2 \leq (1 + \epsilon') \frac{n^3}{12},$$

completing the proof of Theorem 1.3.

Acknowledgment. The author would like to thank Van Vu for very helpful discussions and remarks. He is grateful to the referee for many of his thoughtful comments which led to a substantial improvement in the presentation of the paper.

Appendix A. Proof of Lemma 5.3

Without loss of generality, we assume that m is even. For each $j \leq m/2$, by Cauchy-Schwarz inequality and the triangle inequality, we have

$$\begin{aligned}
\|a_{i_j}\xi\|^2 + \|a + i_{m-j}\xi\|^2 &\geq \frac{1}{2}\|(a + i_{m-j}\xi) - (a + i_j\xi)\|^2 \\
&\geq \frac{1}{2}\|(i_{m-i} - i_j)\xi\|^2 \\
&\geq \frac{1}{2}(i_{m-j} - i_j)^2\|\xi\|^2.
\end{aligned}$$

Summing over $0 \leq j \leq m/2$, we obtain

$$\sum_{j=1}^{m} \|a_{i_j}\xi\|^2 \geq \frac{1}{2} \sum_{1 \leq j \leq m/2} (i_{m-j} - i_j)^2 \|\xi\|^2$$

$$\geq \frac{1}{2} \sum_{1 \leq j \leq m/2} j^2 \|\xi\|^2$$

$$\geq \frac{m^3}{48} \|\xi\|^2,$$

where in the second inequality we used the fact that $i_{m-j} - i_j$ strictly decreases in $\mathbf{Z}$.

Appendix B. Tools from Additive Combinatorics

B.1. **Freiman homomorphism.** We now introduce the concept of a Freiman homomorphism, that allows us to transfer an additive problem in one group G to another group G' in a way which is more flexible than the usual algebraic notion of group homomorphism.

Definition B.2 (Freiman homomorphisms)**.** Let $k \geq 1$, and let X, Y be additive sets of groups G and H respectively. A Freiman homomorphism of order k from X to Y is a map $\phi : X \to Y$ with the property that

$$x_1 + \cdots + x_k = x'_1 + \cdots + x'_k \Longrightarrow \phi(x_1) + \cdots + \phi(x_k) = \phi(x'_1) + \cdots + \phi(x'_k)$$

for all $x_1, \ldots, x_k; x'_1, \ldots, x'_k$. If in addition there is an inverse map ϕ^{-1} from Y to X which is a Freiman homomorphism of order k, then we say that ϕ is a Freiman isomorphism of order k, and that X and Y are Freiman isomorphic of order k.

Clearly Freiman homomorphisms preserve the property of being a progression. We now mention a result that shows torsion-free additive groups are not richer than the integers, for the purposes of understanding sums and differences of finite sets ([14, Chapter 5]).

Theorem B.3. *Let X be a finite subset of a torsion-free additive group G. Then for any integer k, there is a Freiman isomorphism $\phi : X \to \phi(X)$ of order k to some finite subset $\phi(X)$ of the integers $\mathbf{Z}$. The same is true if we replace $\mathbf{Z}$ by $\mathbf{F}_p$, if p is sufficiently large depending on X.*

An identical proof also yields a somewhat stronger result below.

Theorem B.4. *Let X be a finite subset of a torsion-free additive group G. Then for any integer k, there is a map $\phi : X \to \phi(X)$ to some finite subset $\phi(X)$ of the integers $\mathbf{Z}$ such that*

$$x_1 + \cdots + x_i = x_1' + \cdots + x_j' \Leftrightarrow \phi(x_1) + \cdots + \phi(x_i) = \phi(x_1') + \ldots \phi(x_j')$$

for all $i, j \leq k$. The same is true if we replace $\mathbf{Z}$ by $\mathbf{F}_p$, if p is sufficiently large depending on A.

B.5. **Freiman's inverse theorem.** A subset Q of an abelian group is a *Generalized arithmetic progression (GAP) of rank r* if it can be expressed as in the form

$$Q = \{a_0 + x_1 a_1 + \cdots + x_r a_r | M_i \leq x_i \leq M_i' \text{ for all } 1 \leq i \leq r\}$$

for some $a_0, \ldots, a_r$ and $M_1, \ldots, M_r, M_1', \ldots, M_r'$.

It is convenient to think of Q as the image of an integer box $B := \{(x_1, \ldots, x_r) \in \mathbf{Z}^d | M_i \leq m_i \leq M_i'\}$ under the linear map

$$\Phi : (x_1, \ldots, x_d) \mapsto a_0 + x_1 a_1 + \cdots + x_r a_r.$$

The numbers a_i are the *generators* of P, the numbers M_i', M_i are the *dimensions* of P, and $\text{Vol}(Q) := |B|$ is the *volume* of B. We say that Q is *proper* if this map is one to one, or equivalently if $|Q| = \text{Vol}(Q)$. For non-proper GAPs, we of course have $|Q| < \text{Vol}(Q)$. If $-M_i = M_i'$ for all $i \geq 1$ and $a_0 = 0$, we say that Q is *symmetric*.

It is easy to verify that if X is a dense subset of a GAP of bounded rank, then $|X + X|$ has size not much bigger than X. The celebrated Freiman's inverse theorem says the converse. This theorem comes in a number of variants; we give one of them below.

Theorem B.6 (Freiman's inverse theorem). [14, Theorem 5.35] *Let $\gamma \geq 2$ be a given positive number. Let X be a subset of a torsion-free group such that $|X + X| \leq \gamma |X|$. Then for any $0 < \epsilon$ there exists a proper symmetric arithmetic progression Q of rank at most $\lfloor \log_2 \gamma + \epsilon \rfloor$ and size $|Q| = \Theta_{\epsilon, \gamma}(|X|)$ such that $|X \cap Q| = \Theta_{\epsilon, \gamma}(|X|)$.*

In our setting, after passing to $\mathbf{F}_p$ by a Freiman-isomorphism, in order to apply Theorem B.6 we need a to pass back to torsion-free groups. The following result by Green and Ruzsa allows us to do so.

Theorem B.7. [3, Theorem 1.3] *Let $X \subset \mathbf{F}_p$ be a set with $|2X| \leq \gamma |X|$ and $|X| = o_{k,\gamma}(p)$. Then X is Freiman isomorphism of order k to a subset of a torsion-free group.*

B.8. **Sárközy-type theorem in GAP.** Assume that X is a dense subset of a GAP Q, then the iterated sumsets kX contains a structure similar to Q (see [10, Lemma 4.4, Lemma 5.5], [11, Lemma B3]).

Lemma B.9 (Sárközy-type theorem in progressions). *Let Q be a proper GAP in a torsion-free group of rank r. Let $X \subset Q$ be a subset such that $|X| \geq \delta |Q|$ for some $0 < \delta < 1$. Then there exists a positive integer $1 \leq$*

$k \ll_{\delta,r} 1$ *such that kX contains a GAP Q' of rank r and size $\Theta_{\delta,r}(|Q|)$. Furthermore, the generators of Q' are bounded multiples of the generators of Q. If Q and X are symmetric, then Q' can be chosen to be symmetric.*

An easy corollary when Q has rank 1 follows.

Corollary B.10. *Assume that X is a subset of $[-Cn, Cn]$ of size n in $\mathbf{Z}$. Then there exists an integer $k = k(C)$ and a positive number $\gamma = \gamma(C) > 0$ such that $kX - kX$ contains a symmetric arithmetic progression of rank 1 and length $2\gamma n + 1$.*

References

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

[2] P. Erdős and L. Moser, *Elementary Problems and Solutions*: Solutions: E736. Amer. Math. Monthly, 54 (1947), no. 4, 229-230.

[3] B. Green and I. Ruzsa, *Sets with small sumset and rectification*, Bull. London Math. Soc. 38 (2006), no. 1, 43-52.

[4] D. Kleitman, *On a lemma of Littlewood and Offord on the distributions of linear combinations of vectors*, Advances in Math. 5 (1970), 155-157.

[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. 12 , (1943). 277-286.

[6] H. Nguyen and V. Vu, *Optimal Inverse Littlewood-Offord theorems*, Advances in Mathematics, Vol. 226 6 (2011), 5298-5319.

[7] R. A. Proctor, *Solution of two difficult combinatorial problems with linear algebra*. Amer. Math. Monthly 89 (1982), no. 10, 721-734.

[8] A. Sárkőzy and E. Szemerédi, *Über ein Problem von Erdős und Moser*, Acta Arithmetica, 11 (1965) 205-208.

[9] R. Stanley, *Weyl groups, the hard Lefschetz theorem, and the Sperner property*, SIAM J. Algebraic Discrete Methods 1 (1980), no. 2, 168–184.

[10] E. Szemerédi and V. Vu, *Long arithmetic progressions in sumsets: Thresholds and Bounds*. Journal of the A. M. S, 19 (2006), no 1, 119-169.

[11] T. Tao, *Freiman's theorem in solvable groups*, Contributions to Discrete Mathematics, 5 (2010), no. 2, 137-184.

[12] T. Tao and V. Vu, *Inverse Littlewood-Offord theorems and the condition number of random matrices*, Annals of Mathematics (2) 169 (2009), no 2, 595-632.

[13] T. Tao and V. Vu, *On the singularity probability of random Bernoulli matrices*, Journal of the A. M. S 20 (2007), 603-673.

[14] T. Tao and V. Vu, *Additive Combinatorics*, Cambridge Univ. Press, 2006.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF PENNSYLVANIA, PHILADELPHIA, PA 19104

E-mail address: hoing@math.upenn.edu