SOME APPLICATIONS OF FREIMAN'S INVERSE THEOREM

BY HOI H. NGUYEN

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Mathematics

Written under the direction of

Prof. Van Vu

and approved by

New Brunswick, New Jersey

May, 2010

ABSTRACT OF THE DISSERTATION

Some applications of Freiman's inverse theorem

by Hoi H. Nguyen

Dissertation Director: Prof. Van Vu

The celebrated Freiman's inverse theorem in Additive Combinatorics asserts that an additive set of small doubling constant must have additive structure. This thesis contains two applications achieved by combining this theorem with a dyadic pigeonhole principle technique.

1. A finite set A of integers is square-sum-free if no subset of A sums up to a square. In 1986, Erdős posed the problem of determining the largest cardinality of a square-sum-free subset of $\{1, \dots, n\}$.

Significantly improving earlier results, we show in Chapter 2 that this maximum cardinality is of order $n^{1/3+o(1)}$, which is asymptotically tight.

2. A classical result of Littlewood-Offord and Erdős from the 1940s asserts that if the v_i are non-zero, then the concentration probability of the (multi)set $V = \{v_1, \dots, v_n\}$, $\rho(V) := \sup_x \mathbf{P}(v_1\eta_1 + \dots v_n\eta_n = x)$, is of order $O(n^{-1/2})$, where η_i are i.i.d. copies of a Bernoulli random variable.

Motivated by problems concerning random matrices, Tao and Vu introduced the Inverse Littlewood-Offord problem. In the inverse problem, one would like to give a characterization of the set V, given that $\rho(V)$ is relatively large.

In Chapter 3, we develop a method to attack the inverse problem. As an application, we strengthen several previous results of Tao and Vu, obtaining an almost optimal characterization for V. This implies several classical theorems, such as those of Sárközy-Szemerédi, Halász, and Stanley.

Acknowledgements

I would like to thank Prof. Jeff Kahn for his help during my early years at Rutgers University. I am thankful to Prof. Henryk Iwaniec for his suggestion to complete the number theoretic part of Chapter 2.

I am very grateful to Prof. Endre Szemerédi for all the stimulating discussions. Above all, I am deeply indebted to my advisor, Prof. Van Vu, for his enthusiastic encouragement and technical assistances. The main parts of this thesis are joint work with him.

Dedication

This work is dedicated to my parents, ba Thùy, mẹ Tâm, with love.

Table of Contents

Terminology

We use $\mathbf{Z}$ to denote the set of integers, $\mathbf{Q}$ to denote the set of rational numbers, and $\mathbf{F}_p$ to denote the prime field of order p.

Let G be an abelian group. Let A be a subset of G. We denote by S_A the collection of finite partial sums of A,

$$S_A := \left\{ \sum_{x \in B} x; B \subset A, 0 < |B| < \infty \right\}.$$

For two subsets A and B of G, we define their (Minkowski) sum by

$$A + B := \{a + b | a \in A, b \in B\}.$$

For a positive integer l, we define the l-iterated sum of A by

$$lA := \left\{ \sum_{i=1}^{l} a_i | a_i \in A \right\}.$$

For a positive integer $l \leq |A|$ we denote by l^*A the collection of partial sums of l elements of A,

$$l^*A := \left\{ \sum_{x \in B} x; B \subset A, |B| = l \right\}.$$

As usual, $e(x)$ means $\exp(2\pi i x)$, and $e_p(x)$ means $\exp(2\pi i x/p)$.

The notation $[x]$ denotes the set of positive integers at most x.

We use Landau asymptotic notation such as O, Ω, Θ, o under the assumption that $n \to \infty$. Notation such as $\Theta_c(.)$ means that the hidden constant in Θ depends on a (previously defined) quantity c.

We will also omit all unnecessary floors and ceilings. All logarithms have natural base.

Chapter 1

Additive structure and Fremain's inverse theorem

In this chapter we will provide some tools from Additive Combinatorics that will be used for later applications.

1.1 Generalized arithmetic progression (GAP)

A subset Q of an abelian group is a *GAP of rank* r if it can be expressed as in the form

$$Q = \{a_0 + x_1a_1 + \cdots + x_ra_r | M_i \leq x_i \leq M_i' \text{ for all } 1 \leq i \leq r\}$$

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

$$\Phi : (x_1, \ldots, x_d) \mapsto a_0 + x_1a_1 + \cdots + x_ra_r.$$

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

We record a few useful facts about GAPs. Assume that Q is symmetric, $Q = \{a_1x_1 + \ldots a_rx_r : |x_i| \leq M_i, 1 \leq i \leq r\}$. For any $t > 0$, denote by tQ the set

$$\{a_1x_1 + \cdots + a_rx_r : |x_i| \leq tM_i, 1 \leq i \leq r\}.$$

We say that Q is t-proper if tQ is proper. In general, a GAP is not necessarily t-proper. However, one can embed it into a t-proper one with some small loss (see [3, 4], [40, Theorem 3.40]).

Lemma 1.1.1 (Embedding into proper GAP)**.** *Let Q be a symmetric GAP of rank r in a torsion-free group G, and let $t \geq 1$. Then there exists a t-proper symmetric GAP Q' with rank at most r and $|Q'| \leq (2t)^r r^{6r^2}|Q|$ which contains Q. Furthermore, if Q is not proper, we may choose Q' to have rank at most $r-1$.*

Next, assume that A is a dense subset of a GAP Q, then the iterated sumsets kA contains a structure similar to Q (see [30, Lemma 4.4, Lemma 5.5], [32, Lemma B3]).

Lemma 1.1.2 (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 $1 \leq m \ll_{\delta,r} 1$ such that mX 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.*

A more general result holds when one replaces one subset by many subsets of the same GAP.

Lemma 1.1.3 (Sárközy-type theorem in progressions, generalized form)**.** *Let Q be a proper GAP in a torsion-free group of rank r. Let $0 < \delta \leq 1$ be a given constant. Then there exists a positive integer $1 \leq m \ll_{\delta,r} 1$ such that the following holds. If $X_1, \dots, X_m \subset Q$ and $|X_i| \geq \delta|Q|$, then $X_1 + \cdots + X_m$ 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.*

1.2 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 1.2.1 (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 no richer than the integers, for the purposes of understanding sums and differences of finite sets ([40, Chapter 5]).

Theorem 1.2.2. *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.*

By following the same proof, we can show a somewhat stronger result below, which will be used in Chapter 3.

Theorem 1.2.3. *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.

1.3 Freiman's inverse theorem

If X is a dense subset of a GAP, then the doubling constant of X, $\sigma[X] := |2X|/|X|$ is small. The celebrated Freiman's inverse theorem says the converse. This theorem comes in a number of variants; we give two of them below.

Theorem 1.3.1 (Freiman's inverse theorem)**.** *Let γ be a given positive number. Let X be a set in $\mathbf{Z}$ such that $|X+X| \leq \gamma|X|$. Then there exists a proper GAP of rank at most $r = O_\gamma(1)$ and cardinality $O_\gamma(|X|)$ that contains X.*

Freiman's theorem has the following variants ([9, 29], [40, Chapter 5]), which has a weaker conclusion, but provides the optimal estimate for the rank r.

Theorem 1.3.2 (Freiman's inverse theorem)**.** *Let γ, δ be positive constants. Let X be a set in $\mathbf{Z}$ such that $|X+X| \leq \gamma|X|$. Then there exists a proper GAP Q of rank at most $\lfloor \log_2 \gamma + \delta \rfloor$ and cardinality $O_{\gamma,\delta}(|X|)$ such that X is covered by $O_{\gamma,\delta}(1)$ translates of Q.*

We next discuss some crucial results that are directly relevant to our applications.

1.4 Structure in sumsets

One of the most popular problems in Combinatorial Number Theorem is to study whether the iterated sumsets lX of a set X contains a special element (zero, squares, etc) or the whole group. There are various methods to deal with these problems: algebraic, analytic, combinatorial. Basing on the work of Sárközy [26] and Szemerédi-Vu [29, 30], we have developed a new structural approach. In this method, the very first, and most important step, is to find a fine structure in the iterated sumsets. We mention here two such results of Szemerédi and Vu.

Lemma 1.4.1. *For any fixed positive integer d there are positive constants C and c depending on d such that the following holds. For any positive integers n and l and any set X of $[n]$ satisfying $l^d|X| \geq Cn$, lX contains a proper GAP of rank d' and volume at least $cl^{d'}|X|$, for some integer $1 \leq d' \leq d$.*

Lemma 1.4.2. *For any fixed positive integer d there are positive constants C and c depending on d such that the following holds. Let $X_1, \ldots, X_l$ be subsets of $[n]$ of size $|X|$ where l and $|X|$ satisfy $l^d|X| \geq Cn$. Then $X_1 + \cdots + X_l$ contains a GAP of rank d' and volume at least $cl^{d'}|X|$, for some integer $1 \leq d' \leq d$.*

Lemma 1.4.1 and Lemma 1.4.2 will play a key role in Chapter 2.

1.5 Sumsets in structure

In contrast to the previous section, we give here a result showing that in some cases iterated sumsets may be efficiently contained in GAPs ([34, Theorem 1.21]).

Lemma 1.5.1. *Let $A > 0$ be a constant. Assume that X is a subset of integers such that $|lX| \leq l^A|X|$ for some number $l \geq 2$. Then lX is contained in a symmetric 2-proper GAP Q of rank $r = O_A(1)$, and of cardinality $O_A(|lX|)$.*

Using Lemma 1.5.1, we give a structure for X under the condition $|lX| \leq l^A|X|$ in the following theorem, which we will refer to as the Long Range Inverse theorem.

Theorem 1.5.2 (Long Range Inverse theorem)**.** *Let $A > 0$ be constant. Assume that X is a subset of a torsion-free group such that $0 \in X$ and $|lX| \leq l^A|X|$ for some positive integer $l \geq 2$. Then there is proper symmetric GAP Q of rank $r = O(A)$ and cardinality $O_A(l^{-r}|lX|)$ such that $X \subset Q$.*

Notice that for any given $\epsilon > 0$ and if l is large enough, it is implied from Theorem 1.5.2 that the rank of Q is at most $A + \epsilon$. The implicit constant involved in the size of Q can be taken to be $2^{2^{2^{O(A)}}}$, which is quite poor. Although we have not elaborated on this bound much, our method does not seem to say anything when the polynomial growth in size of lX is replaced by something faster.

Theorem 1.5.2 will serve as the main lemma for Chapter 3. To prove it, we combine Lemma 1.5.1 and the following simple observation.

Lemma 1.5.3. *(Dividing sumsets relations) Assume that $0 \in X$ and that $P = \{\sum_{i=1}^{r} x_i a_i : |x_i| \leq N_i\}$ is a symmetric 2-proper GAP that contains lX. Then $X \subset \{\sum_{i=1}^{r} x_i a_i : |x_i| \leq 2N_i/l\}$.*

Chapter 2

Squares in sumsets

2.1 Introduction

In 1986, Erdős [5] raised the following question:

Question 2.1.1. *What is the maximal cardinality of a subset A of $[n]$ such that S_A contains no square?*

We denote by $SF(n)$ the maximal cardinality in question. Erdős observed that

$$SF(n) = \Omega(n^{1/3}). \tag{2.1}$$

To see this, consider the following example

Example 2.1.2. *Let p be a prime and k be the largest integer such that $kp \leq n$. We choose p of order $n^{2/3}$ such that $k = \Omega(n^{1/3})$ and $1 + \cdots + k < p$. Then the set $A := \{p, 2p, \ldots, kp\}$ is square-sum-free.*

Remark 2.1.3. *The fact that p is a prime is not essential. The construction still works if we choose p to be a square-free number, namely, a number of the form $p = p_1 \ldots p_l$ where p_i are different primes.*

Erdős [5] conjectured that $SF(n)$ is close to the lower bound in (2.1). Shortly after Erdős' paper, Alon [1] proved the first non-trivial upper bound

$$SF(n) = O(\frac{n}{\log n}). \tag{2.2}$$

Next, Lipkin [21] improved to

$$SF(n) = O(n^{3/4+o(1)}). \tag{2.3}$$

In [2], Alon and Freiman improved the bound further to

$$SF(n) = O(n^{2/3+o(1)}). \tag{2.4}$$

The latest development was due to Sárközy [26], who showed

$$SF(n) = O(\sqrt{n \log n}). \tag{2.5}$$

In this chapter, we obtain the asymptotically tight bound

$$SF(n) = O(n^{1/3+o(1)}). \tag{2.6}$$

Theorem 2.1.4. *There is a constant C such that for all $n \geq 2$*

$$SF(n) \leq n^{1/3}(\log n)^C \tag{2.7}$$

In fact, we are going to prove the following more general theorem

Theorem 2.1.5. *There is a constant C such that the following holds for all sufficiently large n. Let p be positive integer less than $n^{2/3}(\log n)^{-C}$ and A be a subset of cardinality $n^{1/3}(\log n)^C$ of $[n/p]$. Then there exists an integer z such that $pz^2 \in S_A$.*

Theorem 2.1.4 is the special case when $p = 1$. Furthermore, Theorem 2.1.4 implies many special cases of Theorem 2.1.5. To see this, choose A to have the form $A := \{pb \mid b \in B\}$ where B is a subset of $[n/p]$ and p is a square-free-number. Then finding a square in S_A is the same as finding a number of the form pz^2 in S_B.

If one replaces squares by higher powers, then the problem becomes easier and asymptotic bounds have been obtained earlier (see next section).

2.2 The main ideas

The general strategy for attacking Question 2.1.1 is as follows. One first tries to show that if $|A|$ is sufficiently large, then S_A should contain a large additive structure. Next, one would argue that a large additive structure should contain a square.

In previous works [1, 2, 21, 26], the additive structure was a (homogeneous) arithmetic progression. (An arithmetic progression is homogeneous if it is of the form $\{ld, (l+1)d, \dots, (l+k)d\}$.) It is easy to show that if P is a homogeneous AP of length $C_0 m^{2/3}$ in $[m]$, for some large constant C_0, then P contains a square. Notice that the set S_A is a subset of $[m]$ where $m := |A|n$. Thus, if one can show that S_A contains a homogeneous AP of length $C_0 m^{2/3}$, then we are done. Sárközy could prove that this is indeed the case, given $|A| \geq C_1 \sqrt{n \log n}$ for a properly chosen constant C_1. This also solves (asymptotically) the problem when squares are replaced by higher powers, since in these cases, the lower bound (which can be obtained by modifying Example 2.1.2) is $\Omega(\sqrt{n})$.

Unfortunately, $\sqrt{n}$ is the limit of this argument, since there are examples of a subset A of $[n]$ of size $\Omega(\sqrt{n})$ where the longest AP in S_A is much shorter than $(|A|n)^{2/3}$.

Example 2.2.1. *Consider*

$$A := \{q_1 x_1 + q_2 x_2 | 1 \leq x_i \leq N\}$$

where $q_1 \approx q_2 \approx n^{3/4}$ are different primes and $N = \frac{1}{100} n^{1/4}$. It is easy to show that A is a proper GAP of rank 2 and S_A is contained in the proper GAP

$$\{q_1 x_1 + q_2 x_2 | 1 \leq x_i \leq 1 + \cdots + N\}.$$

Thus, the longest AP in S_A has length at most $1 + \cdots + N = \Theta(n^{1/2})$, while A has cardinality $\Theta(n^{1/2})$.

The key fact that enables us to go below $\sqrt{n}$ and reach the optimal bound $n^{1/3}$ is a recent theorem of Szemerédi and Vu (a special case of Lemma 1.4.1) that showed that if $|A| \geq C n^{1/3}$ for some sufficiently large constant C, then S_A does contain a large proper GAP of rank at most 2.

Lemma 2.2.2. *There are positive constants C and c such that the following holds. If A is a subset of $[n]$ of cardinality at least $Cn^{1/3}$, then S_A contains either an AP Q of length $c|A|^2$ or a proper GAP Q of rank 2 and cardinality at least $c|A|^3$.*

Ideally, the next step would be showing that a large proper GAP Q (which is a subset of $[|A|n]$) contains a square. Thanks to strong tools from number theory, this is not too hard (though not entirely trivial) if Q is homogeneous. However, we do not know how to force this assumption.

The assumption of homogeneity is essential, as without this, one can easily run into local obstructions. For example, if Q is a GAP of the form

$$\{a_0 + a_1x_1 + a_2x_2 | 0 \leq x_i \leq L\}$$

where both a_1 and a_2 are divisible by 6, but $a_0 \equiv 2(\mathrm{mod} 6)$, then clearly Q cannot contain a square, as 2 is not a square modulo 6.

In order to overcome this obstacle, we need to add several twists to the plan. First, we are going to use only a small subset A' of A to create a large GAP Q. Assume that Q has the form

$$\{a_0 + a_1x_1 + a_2x_2 | 0 \leq x_i \leq L\}.$$

(Q can also have rank one but that is the simpler case.) Let q be the g.c.d of a_1 and a_2. If a_0 is a square modulo q, then there is no local obstruction and in principle we can treat Q as if it was homogeneous.

In the next move, we try to add the remaining elements of A (from $A^{''} := A \backslash A'$) to a_0 to make it a square modulo q. This, however, faces another local obstruction. For instance, if in the above example, all elements of $A^{''}$ are divisible by 6, then a_0 will always be $2(\mathrm{mod} 6)$ no matter how we add elements from $A^{''}$ to it.

Now comes a key point. A careful analysis reveals that having all elements of $A^{''}$ divisible by the same integer (larger than one, of course) is the *only* obstruction. Thus, we obtain a useful dichotomy: either S_A contains a square or there is an integer $p > 1$ which is divisible by all elements of a large subset $A^{''}$ of A.

Now we keep working with $A^{''}$. We can write this set as $\{pb \, |b \in B\}$ where B is a subset of $[n/p]$. In order to show that $S_{A^{''}}$ contains a square, it suffices to show that S_B contains a number of the form pz^2. This explains the necessity of Theorem 2.1.5.

A nice feature of the above plan is that it also works for the more general problem considered in Theorem 2.1.5. We are going to iterate, setting new $A := A^{''}$ of the previous step. Since the number of iterations (i.e., the number of p's) is only $O(\log n)$, if we have $|A^{''}| \geq (1 - \frac{1}{(\log n)^c})|A|$ in each step, for a sufficiently large constant c, then the set $A^{''}$ will never be empty and this guarantees that the process should terminate at some point, yielding the desired result.

In the next lemma, which is the main lemma of the chapter, we put these arguments into a quantitative form.

Lemma 2.2.3. *The followings holds for any sufficiently large constant C. Let p be positive integer less than $n^{2/3}(\log n)^{-C}$ and A be a subset of $[n/p]$ of cardinality $n^{1/3}(\log n)^C$. Then there exists $A' \subset A$ of cardinality $|A'| \leq n^{1/3}(\log n)^{C/3}$ such that one of the followings holds (with $A'' := A \backslash A'$)*

- *$S_{A'}$ contains a GAP*
$$Q = \{r + qx \ |0 \leq x \leq L\}$$
where $L \geq n^{2/3}(\log n)^{C/4}$ and $q \leq \frac{n^{2/3}(\log n)^{C/12}}{p}$ and $r \equiv pz^2(\mathrm{mod} q)$ for some integer z.

- *$S_{A'}$ contains a proper GAP*
$$Q = \{r + q(q_1x_1 + q_2x_2) \ |0 \leq x_1 \leq L_1, 0 \leq x_2 \leq L_2, (q_1, q_2) = 1\}$$
such that $\min(L_1, L_2) \geq n^{1/3}(\log n)^{C/4}, L_1L_2 \geq n(\log n)^{C/2}, q \leq \frac{n^{1/3}}{(\log n)^{C/6}p}$ and $r \equiv pz^2(\mathrm{mod} q)$ for some integer z.

- *There exists an integer $d > 1$ such that $d|a$ for all $a \in A''$.*

Given this lemma, we can argue as before and show that after some iterations, one of the first two cases must occur. We show that in these cases the GAP Q should contain a number of the form pz^2, using classical tools from number theory (see Section 2.9 and Section 2.10).

The proof of Lemma 2.2.3 is technical and requires a preparation involving tools from both combinatorics and number theory. These tools will be the focus of the next two sections.

2.3 Further tools from additive combinatorics

Beside Freiman's inverse theorems [1.3.1,1.3.2], we also use the so-called Covering Lemma, due to Ruzsa (see [23],[40, Lemma 2.14]).

Lemma 2.3.1 (Covering Lemma)**.** *Assume that X, Y are finite sets of integers. Then X is covered by at most $|X+Y|/|Y|$ translates of $Y-Y$.*

We say that a GAP $Q = \{a_0 + x_1a_1 + \dots x_da_d | 0 \le x_i \le L_i\}$ is *positive* if its steps a_i's are positive. A useful observation is that if the elements of Q are positive, then Q itself can be brought into a positive form.

Lemma 2.3.2. *A GAP with positive elements can be brought into a positive form.*

Proof. (of Lemma 2.3.2) Assume that

$$Q = \{a_0 + x_1a_1 + \dots x_da_d | 0 \le x_i \le L_i\}.$$

By setting $x_i = 0$, we can conclude that $a_0 > 0$. Without loss of generality, assume that $a_1, \dots, a_j < 0$ and $a_{j+1}, \dots, a_d > 0$. By setting $x_i = 0$ for all $i > j$ and $x_i = L_i, i \le j$, we have

$$a_0' := a_0 + a_1L_1 + \dots a_jL_j > 0.$$

Now we can rewrite Q as

$$Q := \{a_0' + x_1(-a_1) + \dots + x_j(-a_j) + x_{j+1}a_{j+1} + \dots x_da_d | 0 \le x_i \le L_i\},$$

completing the proof. □

Since we only deal with positive integers, this lemma allows us to assume that all GAPs arising in the proof are in positive form.

Using the above tools and ideas from [29], we will prove Lemma 2.3.3 below, which asserts that if a set A of $[n/p]$ is sufficiently dense, then there exists a small set $A' \subset A$ whose subset sums contain a large GAP Q of small rank. Furthermore, the set $A'' =$

$A \backslash A'$ is contained in only a few translates of Q. This lemma will serve as a base from which we will attack Lemma 2.2.3, using number theoretical tools discussed in the next section.

Lemma 2.3.3. *The following holds for all sufficiently large constant C. Let p be positive integer less than $n^{2/3}(\log n)^{-C}$ and A be a subset of $[n/p]$ of cardinality $n^{1/3}(\log n)^{C}$. Then there exists a subset A' of A of cardinality $|A'| \leq n^{1/3}(\log n)^{C/3}$ such that one of the followings holds (with $A'' := A \backslash A'$):*

- *$S_{A'}$ contains an AP*

  $$Q = \{r + qx \; |0 \leq x \leq L\}$$

  where $L \geq n^{2/3}(\log n)^{C/2}$ and there exist $m = O(1)$ different numbers $s_1, \dots, s_m$ such that $A'' \subset \{s_1, \dots, s_m\} + Q$.

- *$S_{A'}$ contains a proper GAP*

  $$Q = \{r + a_1x_1 + a_2x_2) \; |0 \leq x_1 \leq L_1, 0 \leq x_2 \leq L_2$$

  such that $L_1L_2 \geq n(\log n)^{C/2}\}$ and there exists $m = O(1)$ numbers $s_1, \dots, s_m$ such that $A'' \subset \{s_1, \dots, s_m\} + Q$.

Remark 2.3.4. *The proof actually gives a better lower bounds for L_1L_2 in the second case ($2C/3$ instead of $C/2$), but this is not important in applications.*

2.4 Tools from number theory

Fourier Transform and Poisson summation. Let f be a function with support on $\mathbf{Z}$. The Fourier transform $\widehat{f}$ is defined as

$$\widehat{f}(w) := \int_{\mathbf{R}} f(t)e(-wt) \; dt.$$

The classical Poisson summation formula asserts that

$$\sum_{n=-\infty}^{\infty} f(t + nT) = \frac{1}{T} \sum_{m=-\infty}^{\infty} \widehat{f}(\frac{2\pi m}{T})e(mt/T). \tag{2.8}$$

For more details, we refer to [20, Section 4.3].

Smooth indicator functions. We will use the following well-known construction (see for instance [10, Theorem 18] for details).

Lemma 2.4.1. *Let $\delta < 1/16$ be a positive constant and let $[M, M+N]$ be an interval. Then there exists a real function f satisfying the following*

- $0 \leq f(x) \leq 1$ *for any* $x \in \mathbf{R}$.
- $f(x) = 0$ *if* $x \leq M$ *or* $x \geq M+N$.
- $f(x) = 1$ *if* $M + \delta N \leq x \leq M + N(1-\delta)$.
- $|\widehat{f}(\lambda)| \leq 16\widehat{f}(0)\exp(-\delta|\lambda N|^{1/2})$ *for every* λ.

A Weyl type estimate. Next, we need a Weyl type estimate for exponential sums.

Lemma 2.4.2. *For any positive constant ϵ there exist positive constants $\alpha = \alpha(\epsilon)$ and $c(\epsilon)$ such that the following holds. Let a, q be co-prime integers, θ be a real number, and I be an interval of length N. Let M be a positive number such that $MN \geq q^{1+\epsilon}$. Then,*

$$\sum_{\substack{|m|\leq M \\ m\neq 0}} |\sum_{z\in I} e(\frac{amz^2}{q} + \theta mz)| \leq c(M\sqrt{N} + \frac{MN}{\sqrt{q}})(\log MN)^{\alpha}.$$

Quadratic residues. Finally, and most relevant to our problem, we need the following lemma, which shows the existence of integer solutions with given constrains for a quadratic equation.

Lemma 2.4.3. *There is an absolute constants D such that the following holds. Let $a_1, \dots, a_d, r, p, q$ be integers such that $p, q > 0$ and $(a_1, \dots, a_d, q) = 1$. Then the equation*

$$a_1x_1 + \cdots + a_dx_d + r \equiv pz^2 (\mathrm{mod} q) \tag{2.9}$$

has an integer solution $(z, x_1, \dots x_d)$ satisfying $0 \leq x_i \leq (pq)^{1/2}(\log q)^D$.

The rest of the chapter is organized as follows. The proof of the combinatorial statement, Lemma 2.3.3, comes first in Section 2.5. We then start the number theoretical part by giving a proof for Lemma 2.4.2. The verification of Lemma 2.4.3 comes in Section 2.7. After all these preparations, we will be able to establish Lemma 2.2.3 in Section 2.8. The proof of the main result, Theorem 2.1.5, is presented in Sections 2.9 and 2.10.

2.5 Proof of Lemma 2.3.3

We repeat some arguments from [29] with certain modifications. The extra information we want to get here (compared with what have already been done [29]) is the fact that the set $A^{''}$ is covered by only few translates of Q.

2.5.1 An algorithm

Let A' be a subset of cardinality $|A'| = n^{1/3}(\log n)^{C/3}$ and let $A'' := A \backslash A'$. By a simple combinatorial argument (see [29, Lemma 7.9]), we can find in A' disjoint subsets $A'_1, \dots, A'_{m_1}$ such that $|A'_i| \leq 20 \log_2 |A'|$ and $|l_1^* A'_i| \geq |A'|/2$ where

$$l_1 \leq 10 \log_2 |A'| \text{ and } m_1 = |A'|/(40 \log_2 |A'|). \tag{2.10}$$

(For the definition of $l^* A$ see the beginning of the introduction.)

Without loss of generality, we can assume that m_1 is a power of 4. Let $B_1, \dots, B_{m_1}$ be subsets of cardinality $b_1 = |A'|/2$ of the sets $l_1^* A'_1, \dots, l_1^* A'_{m_1}$ respectively. Following [29, Lemma 7.6]), we will run an algorithm with the B_i's as input. The goal of this algorithm is to produce a GAP which has nice relations with $A^{''}$ (while still not as good as the GAP we wanted in the lemma). In the next few paragraphs, we are going to describe this algorithm.

At the first step, set $B_1^1 := B_1, \dots, B_{m_1}^1 := B_{m_1}$ and let $\mathfrak{B}^1 = \{B_1^1, \dots, B_{m_1}^1\}$. Let h be a large constant to be determined later.

At the $(t+1)$-th step, we choose indices i, j and elements $a_1, \dots, a_h \in A''$ that maximizes the cardinality of $\cup_{d=1}^h (B_i^t + B_j^t + a_d)$ (if there are many choices, choose one arbitrarily). Define $B_1^{t+1'}$ to be the union. Delete from A'' the used elements

$a_1, \dots, a_h$, and remove from $\mathfrak{B}^t$ the used sets B_i^t, B_j^t. Find the next maximum union $\cup_{k=1}^h B_i^t + B_j^t + a_k$ with respect to the updated sets $\mathfrak{B}^t$ and A''.

Assume that we have created $m_{t+1} := m_t/4$ sets $B_1^{t+1'}, \dots, B_{m_{t+1}}^{t+1\ '}$. By the algorithm, we have

$$|B_1^{t+1'}| \geq \dots \geq |B_{m_{t+1}}^{t+1\ '}| := b_{t+1}.$$

Now for each $1 \leq i \leq m_{t+1}$ we choose a subset B_i^{t+1} of cardinality exactly b_{t+1} in $B_i^{t+1'}$. These m_{t+1} sets (of the same cardinality) from a collection $\mathfrak{B}^{t+1}$, which is the output of the $(t+1)$-th step.

Since $m_{t+1} = m_t/4$, there are still $m_t/2$ unused sets B_i^t left in $\mathfrak{B}^t$. Without loss of generality, assume that those are $B_1^t, \dots, B_{m_t/2}^t$. With a slight abuse of notation, we use $A^{''}$ at every step, although this set loses a few elements each time. (The number of deleted elements is very small compared to the size of $A^{''}$.)

Let $l_{t+1} := 2l_t + 1$. Observe that

- $l_t \leq 2^t l_1$ (by definition);
- $b_t \leq l_t n/p$ (since $\cup_{d=1}^h (B_i^{t-1} + B_j^{t-1} + a_d) \subset [l_t n/p]$);
- $$|\cup_{d=1}^h B_i^t + B_j^t + a_d| \leq b_{t+1} \tag{2.11}$$

  for all $1 \leq i < j \leq m_t/2$ and $a_1, \dots, a_h \in A''$ (by the algorithm, as it always chooses a union with maximum size).

Now let c be a large constant and k be the largest index such that $b_i \geq cb_{i-1}$ for all $i \leq k$. Then we have

$$c^k b_1 \leq b_k \leq l_k n/p.$$

Since $b_1 = |A'|/2$ and $l_k \leq 2^k l_1$, we deduce an upper bound for k,

$$k \leq \log_{c/2} \frac{l_1 n}{b_1 p}.$$

Next, by the definition of k, we have $b_{k+1} \leq cb_k$. By (2.11), the following holds for all unused sets B_i^k, B_j^k (with $1 \leq i \leq j \leq m_k/2$) and for all $a_1, \ldots, a_h \in A''$:

$$|\cup_{d=1}^{h} (B_i^k + B_j^k + a_d)| \leq b_{k+1} \leq cb_k = c|B_i^k|.$$

In particular

$$|B_1^k + B_i^k| \leq c|B_1^k|$$

holds for all $2 \leq i \leq m_k/2$.

By Plunnecke-Ruzsa estimate (see [40, Corollary 6.28]), we have

$$|B_1^k + B_1^k| \leq c^2|B_1^k|.$$

It then follows from Freiman's theorem, Theorem 1.3.1, that there exists a proper GAP R of rank $O_c(1)$, of size $O_c(1)|B_1^k|$ such that R contains B_1^k. Furthermore, by Lemma 2.3.1, B_i^k is contained in c translates of $B_1^k - B_1^k$, thus B_i^k is also contained in $O_c(1)$ translates of R.

Before continuing, we would like to point out that the parameter h has not yet played any role in the arguments. The freedom of choosing h will be important in what follows. We are going to obtain the desired GAP Q (claimed in the lemma) from R by a few additional operations.

2.5.2 Creation of many similar GAPs.

One problem with R is that its cardinality can be significantly smaller than the bounds on Q in Lemma 2.3.3. We want to obtain larger GAPs by adding many translates of R. While we cannot do exactly this, we can do nearly as good by the following argument, which creates many GAPs which are translates of each other and have cardinalities comparable to that of R.

By the pigeonhole principle, for $i \leq m_k/2$, we can find a set $B_i' \subset B_i^k$ with cardinality $\Theta_c(1)b_k$ which is contained in one translate of R.

By Lemma 1.1.3, there exists $g = O_c(1)$ such that $B'_1 + \cdots + B'_g$ contains a proper GAP Q_1 of cardinality $\Theta_c(1)|R|$. Create Q_2 by summing $B'_{g+1}, \dots, B'_{2g}$, and so on. At the end we obtain $\frac{m_k}{2g} = \Theta_c(1)m_k$ such GAPs. Also, we can require the Q_i's to have the properties below

- rank(Q_i) =rank$(R) = O_c(1)$;
- $|Q_i| = \Theta_c(1)|R| = \Theta_c(1)b_k$;
- each Q_i is a subset of a translate of gR. Thus by Lemma 2.3.1, R is contained in $O_c(1)$ translates of $Q_i - Q_i$;
- the j-th size of Q_i is different from j-th size of R by a (multiplicative) factor of order $\Theta_c(1)$, for all j;
- the j-th step of Q_i is a bounded multiple of the j-th step of R for all j;

Thus, by the pigeonhole principle and truncation (if necessary) we can obtain $m' = \Theta_c(m_k)$ GAPs, say, $Q_1, \dots, Q_{m'}$, which are translate of each other. An important remark here is that since the Q_i are obtained from summing different B's, the sum $Q_1 + \cdots + Q_{m'}$ is a subset of $S_{A'}$. The desired GAP Q will be a subset of this sum.

2.5.3 Embedding A''

In this step, we embed A'' in a union of few translates of a GAP Q_1 of constant rank. We set the (so far untouched) parameter h to be sufficiently large so that

$$\Theta_c(1) = h > c|B_1^k|/|B'_1|.$$

Let d be the largest number such that there are d elements $a_1, \dots, a_d$ of A'' for which the sets $B'_1 + B'_2 + a_i$ are disjoint. Assume for the moment that $d \geq h$, then we would have

$$| \cup_{i=1}^h (B'_1 + B'_2 + a_i) = h|B'_1 + B'_2| \geq h|B'_1| > c|B_1^k|$$

However, this is impossible because $\cup_{i=1}^{h}(B_1'+B_2'+a_i) \subset \cup_{i=1}^{h}(B_1^k+B_2^k+a_i)$ and the latter has cardinality less than $c|B_1^k|$ by definition. Thus we have $d < h$. So $d = O_c(1)$.

Let us fix d elements $a_1, \ldots, a_d$ from A'' which attained the disjointness in the definition of d. By the maximality of d, for any $a \in A''$ there exists a_i so that $(B_1' + B_2' + a) \cap (B_1' + B_2' + a_i) \neq \emptyset$. Hence

$$a - a_i \in B_1^k + B_2^k - (B_1^k + B_2^k) = (B_1^k - B_1^k) + (B_2^k - B_2^k) \subset 2R - 2R.$$

Thus A'' is covered by at most $d = O_c(1)$ translates of $2R - 2R$. On the other hand, since R is contained in $O_c(1)$ translates of $Q_1 - Q_1$, $2R - 2R$ is contained in $O_c(1)$ translates of $4Q_1 - 4Q_1$. It follows that that A'' is covered by $O_c(1)$ translates of Q_1.

The remaining problem here is that Q_1 does not yet have the required rank and cardinality. We will obtain these by adding the Q_i together (recall that these GAPs are translates of each other) and using a rank reduction argument.

2.5.4 Rank reduction

Let P be the symmetric translate of Q_1 (and also of $Q_2, \ldots, Q_{m'}$). Recall that

$$|P| = |Q_1| = \Theta_c(b_k) = \Omega_c(c^k b_1).$$

and also

$$m' = \Theta_c(m_k) = \Theta_c(\frac{b_1}{4^k}), \text{ and } l_{k+1} \leq 2^{k+1} l_1.$$

Set $l := \min\{m', |A'|/2l_{k+1}\}$. Recall that $|A'| = n^{1/3}(\log n)^{C/2}$, $l_1 \leq 10 \log_2 |A'|$ and $b_1 = |A'|/2$. By choosing c and C sufficiently large, we can guarantee that

$$l|P| \geq n^{2/3}(\log n)^{C/2} \ ; l^2|P| \geq n(\log n)^{2C/3}. \tag{2.12}$$

and also

$$l^3|P| \geq n^{4/3}(\log n)^C \tag{2.13}$$

Now we invoke Lemma 1.1.1 to find a large GAP in lP. Assume, without loss of generality, that $l = 2^s$ for some integer s. We start with $P_0 := P$ and $\ell_0 := l$. If $2^s P_0$ is proper, then we stop. If not, then there exists a smallest index i_1 such that $2^{i_1} P_0$ is proper but $2^{i_1+1} P_0$ is not.

By Lemma 1.1.1 (applying to $2^{i_1} P_0$) we can find a symmetric GAP S which contains $2^{i_1} P_0$ such that $rank(S) < r := rank(2^{i_1} P_0)$.

By Lemma 1.1.2, there is a constant $g = \Theta_c(1)$ such that the set $2^g(2^{i_1} P_0)$ contains a symmetric proper GAP P_1 of rank equals $rank(S)$ and cardinality $\Theta_c(1)|2^{i_1} P_0|$. Set $\ell_1 := \ell_0 / 2^{i_1+g}$ if $\ell_0/2^{i_1+g} \geq 1$ and proceed with P_1, ℓ_1 and so on. Otherwise we stop.

Observe that if $2^{i_j} P_j$ is proper, then $|2^{i_j} P_j| = (1 + o(1)) 2^{i_j r_j} |P_j|$, where r_j is the rank of P_j.

As the rank of P_0 is $O_c(1)$, and $r_{j+1} \leq r_j - 1$, we must stop after $\Theta_c(1)$ steps. Let Q' be the symmetric proper GAP Q' obtained when we stop. It has rank d', for some integer $d' < r$ and cardinality at least $\Theta_c(1)\ell_0^{d'}|P_0| = \Theta_c(1) l^{d'} |P|$. On the other hand, since a translate of lP is contained in $S_{A'}$, $|Q'| \leq |A'|n/p \leq |A'|n$, that is $\Theta_c(1) l^{d'}|P| \leq |A'|n$. Because of (2.13), this holds only if $d' \leq 2$.

2.5.5 Properties of Q.

By the Covering Lemma 2.3.1 and by the definition of P_j's, P_i is contained in $O_c(1)$ translates of P_{i+1} for all $i \geq 0$. At the starting point, we know that A'' is contained in $O_c(1)$ translates of P_0. Since there are only $O_c(1)$ different P_j's, at the last step we conclude that $A^{''}$ is covered by $O_c(1)$ translates of Q'.

Furthermore, Q' is a subset of lP. Thus a translate Q of Q' lies in $Q_1 + \cdots + Q_{m'} \subset S_{A'}$. This Q has rank $1 \leq d' \leq 2$ and cardinality $|Q| = |Q'| \geq \Theta(1) l^{d'} |A'|$. (The right hand side satisfies the lower bounds claimed in Lemma 2.3.3, thanks to (2.12).) This is the GAP claimed in Lemma 2.3.3 and our proof is complete.

2.6 Proof of Lemma 2.4.2

If q is a prime, the lemma is a corollary of the well known Weyl's estimate (see [20].) We need to add a few arguments to handle the general case. The following lemma will be useful.

Lemma 2.6.1. *Let $\tau(n)$ be the number of positive divisors of n. For any given $k \geq 3$ there exists a positive constant $\beta(k)$ such that the following holds for every n.*

$$\tau(n) = O_k(\sum_{\substack{d|n \\ d \leq n^{1/k}}} \tau(d)^{\beta(k)}).$$

Proof. (of Lemma 2.6.1). We can set $\beta(k) = k\log(k+1)$. We factorize n in the following specific way

$$n = \prod_{i=1}^{u} p_i^{a_i} \prod_{j=1}^{v} q_j^{b_j}$$

where $p_1 \leq \cdots \leq p_u,\ q_1 \leq \cdots \leq q_v$ are primes and $a_i \geq k > b_j \geq 1$. Set

$$d := \prod_{i=1}^{u} p_i^{\lfloor \frac{a_i}{k} \rfloor} \prod_{j \leq \lfloor \frac{v}{k} \rfloor} q_j.$$

Then $d \leq n^{1/k}$ by definition and

$$(k+1)^k \tau(d)^{\beta(k)} = (k+1)^k 2^{\lfloor \frac{v}{k} \rfloor k\log(k+1)} \prod_{i=1}^{u} (\lfloor \frac{a_i}{k} \rfloor + 1)^{k\log(k+1)} \geq (k+1)^v \prod_{i=1}^{u} (1+a_i) \geq \tau(n),$$

completing the proof. □

Now we start the proof of Lemma 2.4.2. Let $S := \sum_{\substack{|m| \leq M \\ m \neq 0}} |\sum_{z \in I} e(\frac{amz^2}{q} + \theta mz)|$. Following Weyl's argument, we use Cauchy-Schwarz and the triangle inequality to obtain

$$S^2 \leq 2M \sum_{\substack{|m| \leq M \\ m \neq 0}} \sum_{z_1, z_2 \in I} e(\frac{am(z_1 - z_2)(z_1 + z_2)}{q} + \theta m(z_1 - z_2)).$$

For convenience, we change the variables, setting $u := z_1 - z_2, v := z_2$, then

$$S^2 \leq 2M \sum_{\substack{|m|\leq M\\ m\neq 0}} \sum_{|u|\leq N} e(\frac{amu^2}{q} + \theta mu) \sum_{v\in I, v\in I-u} e(\frac{2amuv}{q})$$

$$\leq 2M \sum_{\substack{|m|\leq M\\ m\neq 0}} \sum_{|u|\leq N} |\sum_{v\in I, v\in I-u} e(\frac{2amuv}{q})|.$$

Next, using the basic estimate (see [20, Section 8.2], for instance)

$$|\sum_{K_0<k\leq K_0+K} e(\omega k)| \leq \min(K, \frac{1}{\|2\omega\|})$$

we obtain that

$$S^2 \leq 2M \sum_{\substack{|m|\leq M\\ m\neq 0}} \sum_{|u|\leq N} \min(N, \frac{1}{\|2amu/q\|}).$$

To estimate the right hand side, let N_r be the number of pairs (m,u) such that $2amu \equiv r(\text{mod} q)$. (In what follows, it is useful to keep in mind that a and q are co-primes.) We have

$$S(M,N,q)^2 \leq 2M \left(N_0 N + \sum_{1\leq r\leq q/2} (N_r + N_{q-r})\frac{q}{r} \right). \tag{2.14}$$

To finish the proof, we are going to derive a (uniform) bound for the N_r's. For $0 \leq r \leq q-1$ let $0 \leq r_a \leq q-1$ be the only number such that $ar_a \equiv r(\text{mod} q)$. Thus $2amu \equiv r(\text{mod} q)$ is equivalent with $2mu \equiv r_a(\text{mod} q)$.

First we consider the case $r \neq 0$, thus $r_a \neq 0$. Write $2mu = r_a + sq$. It is clear that $r_a + sq \neq 0$ for all s. Since $2mu \leq 2MN$, we have $|s| \leq 2MN/q$. For each given s the number of such pairs (m,u) is bounded by $\tau(r_a + sq)$.

Choose $k = \max(\frac{1}{\epsilon} + 2, 3)$, then $MN/q \geq (MN)^{2/k}$ by the assumption $MN \geq q^{1+\epsilon}$. It follows from Lemma 2.6.1 that, for $r \neq 0$,

$$\begin{aligned} N_r \leq \sum_{|s|\leq 2MN/q} \tau(r_a+sq) &= O_\epsilon(\sum_{d\leq (MN)^{1/k}} \tau(d)^{\beta(k)}(\sum_{\substack{|s|\leq 4MN/q \\ d|r_a+sq}} 1) \\ &= O_\epsilon(\sum_{d\leq (MN)^{1/k}} \tau(d)^{\beta(k)}(\frac{4MN}{qd}+O(1))) \\ &= O_\epsilon(\frac{MN}{q}\sum_{d\leq (MN)^{1/k}} \frac{\tau(d)^{\beta(k)}}{d}+O((MN)^{1/k+o(1)})) \\ &= O_\epsilon(\frac{MN}{q}\sum_{d\leq (MN)^{1/k}} \frac{\tau(d)^{\beta(k)}}{d}). \end{aligned}$$

Notice that $\sum_{d\leq x}\tau(d)^{\beta(k)} \ll x\log^{\beta'(k)} x$ for some positive constant $\beta'(k)$ depending on $\beta(k)$ (see [20, Section 1.6], for instance). By summation by parts we deduce that

$$N_r = O_\epsilon(\frac{MN}{q}\log^{\beta''(k)}(MN))$$

for some positive constant $\beta''(k)$ depending on $\beta'(k)$.

Now we consider the case $r=0$. The equation $2mu = sq$ has at most $\tau(sq)$ solution pairs (m,u), except when $s=0$, the case that has $2M$ solutions $\{(m,0); |m|\leq 2M, m\neq 0\}$. Thus we have

$$N_0 \leq 2M + \sum_{|s|\leq 2MN/q, s\neq 0} \tau(sq),$$

and hence,

$$N_0 = O_\epsilon(2M+\frac{MN}{q}\log^{\beta''(k)}(MN)).$$

Combining these estimates with (2.14), we can conclude that

$$S(M,N,q) \ll_\epsilon (M\sqrt{N}+MN/\sqrt{q})\log^\alpha(MN)$$

for some sufficiently large constant $\alpha = \alpha(\epsilon)$.

2.7 Proof of Lemma 2.4.3

We are going to need the following simple fact.

Fact 2.7.1. *Let $a_1, \dots, a_m, q$ be integers such that $(a_1, \dots, a_m, q) = 1$. Then we can select a decomposition $q = q_1 \dots q_l$ of q and l different numbers $a_{i_1}, \dots, a_{i_l}$ of $\{a_1, \dots, a_m\}$ (for some $l \geq 1$) such that*

$$(q_i, q_j) = 1 \textit{ for evey } i \neq j \textit{ and } (a_{i_j}, q_j) = 1 \textit{ for every } j.$$

Proof. (of Fact 2.7.1) Let $q = q'_1 \dots q'_k$ be the decomposition of q into prime powers. For each q'_i we assign a number a'_i from $\{a_1, \dots, a_m\}$ such that $(q'_i, a'_i) = 1$ (the same a_i may be assigned to many q'_j). Let a_{i_j}'s be the collection of the a'_i's without multiplicity. Set q_j to be the product of all q'_i assigned to a_{i_j}. □

The core of the proof of Lemma 2.4.3 will be the following proposition, which is basically the case of one variable in a slightly more general setting.

Proposition 2.7.2. *There is a constants D such that the following holds. For given integers $g, h, p, t, z_1; g, h, p > 0$ there exist integers $x \in [0, (ph)^{1/2}(\log h)^D]$ and z_2 such that $gx + pz_1^2 + tk \equiv pz_2^2 (\mathrm{mod} h)$, where $k = (g, h)$.*

Lemma 2.4.3 follows from Fact 2.7.1 and Proposition 2.7.2 by an inductive argument. Indeed, by the above fact we may assume that $q = q_1 \dots q_l$ where $(a_i, q_i) = 1$, and so

$$(a_l, q) | q_1 \dots q_{l-1}.$$

Now if Lemma 2.4.3 is true for $l - 1$ variables, i.e. there are appropriate $x_1, \dots, x_{l-1}$ such that $a_1 x_1 + \dots a_{l-1} x_{l-1} + r = pz_1^2 + tq_1 \dots q_{l-1}$. Then we apply Proposition 2.7.2 for $q = h, g = a_l$ to find x_l. It thus remains to justify Proposition 2.7.2.

Proof. (of Proposition 2.7.2) Without loss of generality we assume that $h \geq 3$. As $k = (g, h)$, we can write $g = ka, h = kq$ where $(a, q) = 1$. We shall find a solution in the form $z_2 = z_1 + zk$. Plugging in z_2 in this form and simplifying by k, we end up with the equation

$$ax + t \equiv pkz^2 + 2pz_1z(\text{mod}q).$$

or equivalently,

$$x \equiv \bar{a}pkz^2 + 2\bar{a}pz_1z - \bar{a}t(\text{mod}q) \tag{2.15}$$

where $\bar{a}$ is the reciprocal of a modulo q, $a\bar{a} \equiv 1(\text{mod}q)$.

Our task is to find $x \in [0, (ph)^{1/2}(\log h)^D]$ such that (2.15) holds for some integer z. Notice that if q is small and D is large then $(ph)^{1/2}(\log h)^D \geq (\log 3)^D$, therefore the interval $[0, (ph)^{1/2}]$ contains every residue class modulo q; as a result, (2.15) holds trivially. From now on we can assume that q is large,

$$q \geq \exp\left(16(6(\alpha+1)/e)^{\alpha+1}\right) \tag{2.16}$$

where c, α are constants arising from Lemma 2.4.2 with $\epsilon = 1/3$.

Let $s = (pk, q)$; so we can write $pk = sp', q = sq'$ with $(p', q') = 1$.

Let D be a large constant (to be determined later) and set

$$L := (sq)^{1/2}(\log q)^D/2 \text{ and } I := [L, 2L].$$

Note that

$$ph = pkq = sp'q \geq sq.$$

Thus we have

$$I \subset [0, (ph)^{1/2}(\log h)^D].$$

Let f be a smooth function defined with respect to the interval I (as in Lemma 2.4.1). For fixed $z \in [1, q]$ the numbers of x in $[0, (sq)^{1/2}\log^D q]$ satisfying (2.15) is at least

$$N_z := \sum_{m \in \mathbf{Z}} f(\bar{a}pkz^2 + 2\bar{a}pz_1z - \bar{a}t + mq).$$

By Poisson summation formula (2.8)

$$N_z = \sum_{m\in\mathbf{Z}} \frac{1}{q}\widehat{f}(\frac{m}{q})e(\frac{(\bar{a}pkz^2+2\bar{a}pz_1z-\bar{a}t)m}{q}).$$

By summing over $z \in [1,q]$ we obtain

$$N := \sum_{z=1}^{q} N_z = \frac{1}{q}\sum_{m\in\mathbf{Z}} \widehat{f}(\frac{m}{q})\sum_{z=1}^{q} e(\frac{(\bar{a}pkz^2+2\bar{a}pz_1z-\bar{a}t)m}{q}).$$

To conclude the proof, it suffices to show that $N > 0$. We are going to show (as fairly standard in this area) that the sum is dominated by the contribution of the zero term. By the triangle inequality, we have

$$|N-\widehat{f}(0)| \leq \frac{1}{q}\sum_{m\in\mathbf{Z},m\neq 0} |\widehat{f}(\frac{m}{q})||\sum_{z=1}^{q} e(\frac{(\bar{a}pkz^2+2\bar{a}pz_1z)m}{q})|.$$

Let γ_1, γ_2 be a sufficiently large constant and let

$$L' := \frac{\gamma_1 q(\log q)^{\gamma_2}}{L}.$$

Set

$$S_1 := \frac{1}{q}\sum_{|m|\geq L'} |\widehat{f}(\frac{m}{q})||\sum_{z=1}^{q} e(\frac{(\bar{a}pkz^2+2\bar{a}pz_1z)m}{q})|$$

and

$$S_2 := \frac{1}{q}\sum_{\substack{|m|\leq L' \\ m\neq 0}} |\widehat{f}(\frac{m}{q})||\sum_{z=1}^{q} e(\frac{(\bar{a}pkz^2+2\bar{a}pz_1z)m}{q})|.$$

We then have

$$|N-\widehat{f}(0)| \leq S_1 + S_2.$$

In what follows, we show that both S_1 and S_2 are less than $\widehat{f}(0)/4$.

Estimate for S_1. It is not hard to show that

$$\sum_{k\in\mathbf{Z}} \exp(-\sqrt{x|k|}) < \frac{5}{x} \text{ for } 0 < x < 1.$$

To see this, observe that

$$\sum_{k\geq 1} \exp(-\sqrt{xk}) \leq \int_0^\infty \exp(-\sqrt{xt})dt = \frac{2}{x},$$

where the integral is evaluated by changing variable and integration by parts.

Thus

$$\sum_{|k|\geq k_0} \exp(-\sqrt{x|k|}) < \sum_{k\in\mathbf{Z}} \exp(-\sqrt{x}(\frac{\sqrt{|k|}+\sqrt{k_0}}{2})) \leq \frac{20}{x}\exp(-\frac{\sqrt{xk_0}}{2}). \qquad (2.17)$$

From the property of f (Lemma 2.4.1) we can deduce that

$$S_1 \leq 16\widehat{f}(0) \sum_{|m|\geq \frac{\gamma_1 q(\log q)^{\gamma_2}}{L}} \exp(-\delta\sqrt{|Lm/q|}),$$

which, via (2.17) and since $q \geq 3$, implies

$$S_1 \leq 16\widehat{f}(0)\frac{20}{Lq^{-1}}\exp(-\frac{\delta(\gamma_1(\log q)^{\gamma_2})^{1/2}}{2}) \leq \widehat{f}(0)/4,$$

given that we choose γ_1, γ_2 sufficiently large.

Estimate for S_2. We have

$$S_2 = \frac{\widehat{f}(0)}{q} \sum_{\substack{|m|\leq L' \\ m\neq 0}} |\sum_{z=1}^{q} e(\frac{\bar{a}p'z^2}{q'} + \frac{2\bar{a}pz_1zm}{q})|.$$

We shall choose $D > \gamma_2$.

Set

$$\gamma_1 := (\frac{6(D-\gamma_2)}{e})^{D-\gamma_2}.$$

First, we observe that

$$L'q = \frac{2\gamma_1 q^2(\log q)^{\gamma_2}}{(sq)^{1/2}(\log q)^D} = \frac{2\gamma_1 q^{3/2}}{s^{1/2}(\log q)^{D-\gamma_2}} = \frac{2\gamma_1 {q'}^{1/2} q}{(\log q)^{D-\gamma_2}} \geq {q'}^{4/3} \frac{\gamma_1 q^{1/6}}{(\log q)^{D-\gamma_2}}.$$

It is not hard to show that the function $q^{1/6}/(\log q)^{D-\gamma_2}$, where $q \geq 3$, attains its minimum at $q = \exp(6(D-\gamma_2))$. Therefore, by the choice of γ_1, we have

$$L'q \geq {q'}^{4/3}.$$

Next, Lemma 2.4.2 applied for $\epsilon = 1/3$ (and with the mentioned c and α) yields

$$\begin{aligned} S_2 &= \frac{\widehat{f}(0)}{q} \sum_{\substack{|m| \leq L' \\ m \neq 0}} |\sum_{z=1}^{q} e(\frac{\bar{a}p'z^2}{q'} + \frac{2\bar{a}pz_1zm}{q})| \\ &\leq c\frac{\widehat{f}(0)}{q}(\frac{L'q}{\sqrt{q'}} + L'\sqrt{q})(\log q)^\alpha \\ &\leq 2c\frac{\widehat{f}(0)}{q}\frac{L'q}{\sqrt{q'}}(\log q)^\alpha = 2c\frac{\widehat{f}(0)L'}{\sqrt{q'}}(\log q)^\alpha. \end{aligned}$$

It follows that

$$S_2 \leq \frac{4c\gamma_1 q(\log q)^{\alpha+\gamma_2}}{(\sqrt{sq}\log^D q)\sqrt{q'}}\widehat{f}(0) = \frac{4c\gamma_1(\log q)^{\alpha+\gamma_2}}{(\log q)^D}\widehat{f}(0).$$

Now we choose D, γ_2 so that $D - \gamma_2 - \alpha = 1$. Thus $\gamma_1 = (6(\alpha+1)/e)^{\alpha+1}$, and

$$S_2 \leq \frac{4c\gamma_1(\log q)^{\alpha+\gamma_2}}{(\log q)^D}\widehat{f}(0) = \frac{4c(6(\alpha+1)/e)^{\alpha+1}}{\log q}\widehat{f}(0) \leq \widehat{f}(0)/4$$

where the last inequality comes from (2.16).

□

2.8 Proof of Lemma 2.2.3

We first apply Lemma 2.3.3 to obtain a large proper GAP Q of rank 1 or 2. By this lemma, we have $A'' \subset \{s_1, \ldots, s_m\} + Q$, where m is a constant.

Let $S_i = A'' \cap (s_i + Q)$ for $1 \leq i \leq m$. We would like to guarantee that all S_i are large by the following argument.

If S_i is smaller than $n^{1/3}(\log n)^{3C/10}$, then we delete it from $A^{''}$ and add to A'. The new sets A', $A^{''}$ and Q still satisfy the claim of Lemma 2.3.3. On the other hand, that the total number of elements added to A' is only $O(n^{1/3}(\log n)^{3C/10} = o(|A'|)$, thus the sizes of A' and A'' hardly changes.

From now on, we assume that $|S_i| \geq n^{1/3}(\log n)^{3C/10}$ for all i.

For convenience, we let

$$s_i' := s_i + r.$$

Thus every element of S_i is congruent with s_i' modulo q.

2.8.1 Q has rank one

In this subsection, we deal with the (easy) case when Q has rank one. We write $Q = \{r + qx \; |0 \leq x \leq L\}$ where $L \geq n^{2/3}(\log n)^{C/2}$.

Since $Q \subset S_{A'} \subset [\frac{n}{p}|A'|]$, we have

$$q \leq \frac{|A'|n}{pL} \leq \frac{n^{2/3}}{(\log n)^{C/6}p}.$$

By setting C (of Lemma 2.3.3) sufficiently large compared to D (of Lemma 2.4.3), we can guarantee that

$$(pq)^{1/2}(\log q)^D \leq n^{1/3}. \tag{2.18}$$

Let $d := (s_1 + r, \ldots, s_m + r, q) = (s_1', \ldots, s_m', q)$. If $d > 1$ then all elements of A'' are divisible by d, since $A^{''}$ are covered by $\{s_1, \ldots, s_m\} + Q$. Thus we reach the third case of the lemma and are done.

Assume now that $d = 1$. By Lemma 2.4.3, we can find $0 \leq x_i \leq (pq)^{1/2}(\log q)^D$ such that

$$s_1'x_1 + \cdots + s_m'x_m + r \equiv pz^2(\text{mod} q). \tag{2.19}$$

Pick from S_i's exactly x_i elements and add them together to obtain a number s. The set $s+Q$ is a translate of Q which satisfies the first case of Lemma 2.2.3 and we are done.

2.8.2 Q has rank two

In this section, we consider the (harder) case when Q has rank two. The main idea is similar to the rank one case, but the technical details are somewhat more tedious. We write

$$Q = r + q(q_1x + q_2y)|0 \leq x \leq L_1, 0 \leq y \leq L_2$$

where $L_1L_2 = |Q| \geq n \log^{2C/3} n$.

As Q is proper, either $q_1 \geq L_2$ or $q_2 \geq L_1$ holds. Thus $qL_1L_2 \leq |A'|n/p$, which yields (with room to spare)

$$q \leq \frac{n^{1/3}}{(\log n)^{C/6}p}. \tag{2.20}$$

We consider two cases. In the first (simple) case, both L_1 and L_2 are large. In the second, one of them can be small.

Case 1. $\min(L_1, L_2) \geq n^{1/3}(\log n)^{C/4}$. Define $d := (s'_1, \dots, s'_m, q)$ and argue as in the previous section. If $d > 1$, then we end up with the third case of Lemma 2.2.3. If $d = 1$ then apply Lemma 2.4.3. The fact that q is sufficiently small (see (2.20)) and that $|S_i|$ is sufficiently large guarantee that we can choose x_i elements from S_i. At the end, we will obtain a GAP of rank 2 which is a translate of Q and satisfies the second case of Lemma 2.2.3.

Case 2. $\min(L_1, L_2) \leq n^{1/3}(\log n)^{C/4}$. In this case the sides of GAP Q are unbalanced and one of them is much larger than the other. We are going to exploit this fact to create a GAP of rank one (i.e., an arithmetic progression) which satisfies the first case of Lemma 2.3.3, rather than trying to create a GAP of rank two as in the previous case.

Without loss of generality, we assume that $L_1 \leq n^{1/3}(\log n)^{C/4}$. By the lower bound on L_1L_2, we have that $L_2 \geq n^{2/3}(\log n)^{C/4}$. This implies

$$qq_2 \leq \frac{|A'|n}{pL_2} \leq \frac{n^{2/3}}{(\log n)^{C/12}p}.$$

Again by setting C sufficiently large compared to D, we have

$$(pqq_2)^{1/2}(\log qq_2)^D \leq n^{1/3}(\log n)^{C/5}. \tag{2.21}$$

Creating a long arithmetic progression. In the rest of the proof we make use of A'' and Q to create an AP of type $\{r' + qq_2x_2 \,|0 \leq x_2 \leq L_2, r' \equiv pz^2(\text{mod} qq_2)\}$. This gives the first case in Lemma 2.3.3 and thus completes the proof of this lemma.

Let S be an element of $\{S_1, \dots, S_m\}$. Since S is contained in a translate of Q, there is a number s such that any $a \in S$ satisfies $a \equiv s + tqq_1(\text{mod} qq_2)$ for some $0 \leq t \leq L_1$ (for instance, if $a \in S_i$ then $a \equiv s'_i + tqq_1(\text{mod} qq_2)$). Let T denote the *multiset* of t's obtained this way. Notice that T could contain one element of multiplicity $|S|$. Also recall that $|S| \geq n^{1/3}(\log n)^{3C/10}$.

For $0 \leq l \leq |S|/2$, let m_l and M_l (respectively) be the minimal and maximal values of the sum of l elements of T. Since $0 \leq t \leq L_1$ for every $t \in T$, by swapping summands of m_l with those of M_l, we can obtain a sequence $m_l = n_0 \leq \dots \leq n_l = M_l$ where each $n_i \in l^*T$ and $n_{i+1} - n_i \leq L_1$ for all relevant i.

By construction, we have

$$[m_l, M_l] \subset \{n_0, \dots, n_l\} + [0, L_1] \subset l^*T + [0, L_1]. \tag{2.22}$$

Next we observe that if l is large and $M_l - m_l$ is small, then T looks like a sequence of only one element with high multiplicity. We will call this element the *essential* element of T.

Proposition 2.8.3. *Assume that* $\frac{1}{4}(n^{1/3}(\log n)^{3C/10} \leq l \leq \frac{1}{2}n^{1/3}(\log n)^{3C/10}$ *and* $M_l - m_l < \frac{1}{4}n^{1/3}(\log n)^{3C/10}$. *Then all but at most* $\frac{1}{2}n^{1/3}(\log n)^{3C/10}$ *elements of* T *are the same.*

Proof. (of Proposition 2.8.3) Let $t_1 \leq t_2 \leq \dots \leq t_l$ be the l smallest elements of T and $t'_1 \leq \dots \leq t'_l$ be the l largest. By the upper bound on l and lower bound on

$|S| = |T|$, $t'_1 \geq t_l$. On the other hand, $M_l - m_l = (t'_1 - t_1) + \cdots + (t'_l - t_l)$. Thus if $M_l - m_l < \frac{1}{4}n^{1/3}(\log n)^{3C/10} \leq l - 1$ then $t'_i = t_i$ for some i. The claim follows. □

The above arguments work for any S among $S_1, \ldots, S_m$. We now associate to each S_i a multiset T_i, for all $1 \leq i \leq m$.

Subcase 2.1 *The hypothesis in Proposition 2.8.3 holds for all T_i.* In this case we move to A' those elements of S_i whose corresponding parts in T_i is not the essential element. The number of elements moved is only $O(n^{1/3}(\log n)^{3C/10})$, which is negligible compared to both $|A'|$ and $|A''|$. Furthermore, the properties claimed in Lemma 2.3.3 remain unchanged and the size of new S_i are now at least $\frac{1}{2}n^{1/3}(\log n)^{3C/10}$.

Now consider the elements of A'' with respect to modulo qq_2. Since each T_i has only the essential element, the elements of $A^{''}$ produces at most m residues $u_i = s'_i + t_i qq_1$, each of multiplicity at least

$$|S_i| \geq \frac{1}{2}n^{1/3}(\log n)^{3C/10} \geq (pqq_2)^{1/2}(\log qq_2)^D$$

where the last inequality comes from (2.21). Define $d = (u_1, \ldots, u_m, qq_2)$ and proceed as usual, applying Lemma 2.4.3.

Subcase 2.2 *The hypothesis in Proposition 2.8.3 does not hold for all T_i.* We can assume that, with respect to T_1, $M_l - m_l \geq \frac{1}{4}n^{1/3}(\log n)^{3C/10}$ for all $\frac{1}{4}n^{1/3}(\log n)^{3C/10} \leq l \leq \frac{1}{2}n^{1/3}(\log n)^{3C/10}$. From now on, fix an l in this interval.

Next, for a technical reason, we extract from S_1 a very small part S'_1 of cardinality $n^{1/3}(\log n)^{C/5}$ and set $S^{''}_1 = S_1 \backslash S'_1$. Let T be the multiset associated with $S^{''}_1$. We can assume that T satisfies the hypothesis of this subcase.

Define $d := (s'_1, \ldots, s'_m, q)$. As usual, the case $d > 1$ leads to the third case of Lemma 2.2.3, so we can assume $d = 1$. By Lemma 2.4.3, there exist integers

$$0 \leq x_i \leq (pq)^{1/2}(\log n)^D \leq n^{1/3}(\log n)^{C/5} \leq |S_i|$$

and k, z_1 such that

$$s'_1 x_1 + \cdots + s'_m x_m + (ls'_1 + r) = pz_1^2 + kq. \tag{2.23}$$

For $i \geq 2$ we pick from S_i exactly x_i elements $a^i_1, \dots, a^i_{x_i}$, and for $i = 1$ we pick x_1 elements $a^1_1, \dots, a^1_{x_1}$ from S'_1 and add them together. By (2.23) the following holds for some integer k',

$$\sum_{i=1}^{m} \sum_{j=1}^{x_i} a^i_j + (ls'_1 + r) = pz_1^2 + k'q. \tag{2.24}$$

Furthermore, by Proposition 2.7.2, as $q = (qq_1, qq_2)$, there exist $0 \leq x \leq (pqq_2)^{1/2} \log^D(qq_2)$ and k'', z_2 such that

$$qq_1x + pz_1^2 + (k' + m_lq_1)q = pz_2^2 + k''qq_2,$$

$$pz_1^2 + k'q + (x + m_l)qq_1 = pz_2^2 + k''qq_2. \tag{2.25}$$

As $(pqq_2)^{1/2} \log^D(qq_2) \leq n^{1/3} \log^{C/5} n$ and $n^{1/3} \log^{C/5} n \leq M_l - m_l$, we have

$$m_l \leq x + m_l \leq M_l.$$

On the other hand, recall that $[m_l, M_l] \subset l^*T + [0, L_1]$ (see (2.22)), we have

$$\{ls'_1 + r + [m_l, M_l]qq_1\} \subset l^*S''_1 + r + [0, L_1]qq_1(\mathrm{mod} qq_2).$$

Thus

$$ls'_1 + r + (x + m_l)qq_1 \in l^*S''_1 + r + [0, L_1]qq_1(\mathrm{mod} qq_2). \tag{2.26}$$

Combining (2.24),(2.25) and (2.26) we infer that there exist l elements $a_1, \dots, a_l$ of S''_1, and there exist $0 \leq u \leq L_1$ and v such that

$$\sum_{i=1}^{m} \sum_{j=1}^{x_i} a^i_j + a_1 + \cdots + a_l + r + uqq_1 = pz_2^2 + vqq_2.$$

Hence, $\sum_{i=1}^m \sum_{j=1}^{x_i} a^i_j + a_1 + \cdots + a_l + Q$ contains the AP $\{(pz_2^2 + vqq_2) + qq_2x_2 | 0 \leq x_2 \leq L_2\}$, completing Subcase 2.2.

Finally, one checks easily that the number of elements of A'' involved in the creation of pz_2^2 in all cases is bounded by $O(n^{1/3}\log^{C/5} n) = o(|A'|)$, thus we may put all of them to A' without loss of generality.

2.9 Proof of Theorem 2.1.5: The rank one case.

Here we consider the (easy) case when Q (in Lemma 2.2.3) has rank one. In this case, $S_{A'}$ contains an AP $Q = \{r + qx | 0 \leq x \leq L\}$, where $L \geq n^{2/3}(\log n)^{C/4}$ as in the first statement of Lemma 2.2.3. We are going to show that Q contains a number of the form pz^2.

Write $r = pz_0^2 + tq$ for some $0 \leq z_0 \leq q$. Since r is a sum of some elements of A', we have

$$0 \leq r \leq |A'|(n/p) \leq \frac{n^{4/3}(\log n)^{C/3}}{p}.$$

Thus

$$-pq \leq t \leq \frac{n^{4/3}(\log n)^{C/3}}{pq}. \tag{2.27}$$

The interval $[t/pq, (t+L)/pq]$ contains at least two squares because

$$(\frac{L}{pq})^2 \geq \frac{n^{4/3}(\log n)^{C/2}}{(pq)^2} \geq 10\frac{t}{pq} + 20.$$

Thus, we can find an integer $x_0 \geq 0$ such that $\frac{t}{pq} < x_0^2 < (x_0+1)^2 \leq \frac{t+L}{pq}$. It is implied that (since $0 \leq z_0 \leq q$)

$$t \leq pqx_0^2 + 2pz_0x_0 \leq t + L. \tag{2.28}$$

Set $z := z_0 + qx_0$. We have

$$pz^2 = pz_0^2 + q(pqx_0^2 + 2pz_0x_0).$$

On the other hand, by (2.28), the right hand side belongs to

$$pz_0^2 + q[t, t+L] = pz_0^2 + tq + q[0, L] = r + q[0, L] = Q.$$

Thus, Q contains pz^2, completing the proof for this case.

2.10 Proof of Theorem 2.1.5: The rank two case

In this case, we assume that $S_{A'}$ contains a proper GAP as in the second statement of Lemma 2.2.3. We can write

$$Q = \{r + q(q_1x_1 + q_2x_2) \; |0 \le x_1 \le L_1, 0 \le x_2 \le L_2, (q_1, q_2) = 1\}$$

where

- $\min(L_1, L_2) \ge n^{1/3}(\log n)^{C/4}$,
- $L_1L_2 \ge n(\log n)^{C/2}$,
- $q \le \frac{n^{1/3}(\log n)^{-C/6}}{p}$,
- and $r = pz_0^2 + tq$ for some integers t and $0 \le z_0 \le q$.

Since r is a sum of some elements of A', we have $0 \le r \le \frac{n^{4/3}(\log n)^{C/3}}{p}$, and so

$$-pq \le t \le \frac{n^{4/3}(\log n)^{C/3}}{pq}.$$

Without loss of generality, we assume that $q_2L_2 \ge q_1L_1$. Because Q is proper, either $q_2 \ge L_1$ or $q_1 \ge L_2$. On the other hand, if $q_2 < L_1$ then $L_2 \le q_1$, which is impossible by the assumption. Hence,

$$q_2 \ge L_1.$$

Now we write $Q = \{pz_0^2 + q(q_1x_1 + q_2x_2 + t)|0 \le x_1 \le L_1, 0 \le x_2 \le L_2, (q_1, q_2) = 1\}$ and notice that if we set $w := z_0 + zq$ then

$$pw^2 - pz_0^2 = p(z_0 + qz)^2 - pz_0^2 = q(pqz^2 + 2pz_0z).$$

Thus if there is an integer z satisfies

$$pqz^2 + 2pz_0z \in \{q_1x + q_2y + t|0 \leq x \leq L_1, 0 \leq y \leq L_2\} \tag{2.29}$$

then $pw^2 \in Q$, and we are done with this case. The rest of the proof is the verification of the following proposition, which shows the existence of a desired z.

Proposition 2.10.1. *There exists an integer z which satisfies (2.29).*

Proof. (of Proposition 2.10.1) The method is similar to that of Lemma 2.4.3, relying on Poisson summation.

Set $a := pq$ and $b := 2pz_0$. Notice that since $0 \leq z_0 \leq q$, $0 \leq b \leq 2pq = 2a$. Our task is to find a z such that

$$az^2 + bz - q_1x - t = q_2y \text{ for some } 0 \leq x \leq L_1, 0 \leq y \leq L_2.$$

Define (with foresight; see (2.31)) $I_x := [L_1/8, L_1/4]$ and

$$I_z := [(\frac{q_1L_1/4 + t}{a})^{1/2} + 1, (\frac{q_2L_2 + q_1L_1/8 + t}{a})^{1/2} - 1].$$

(Notice the that the lower bounds on L_1, L_2 and the upper bound on pq guarantee that the expressions under the square roots are positive.)

Since $r + qq_1L_1 + qq_2L_2 = pz_0^2 + tq + q(q_1L_1 + q_2L_2) \in Q$, it follows that (with $\max(Q)$ denoting the value of the largest element of Q)

$$q_2L_2 + q_1L_1/8 + t \leq \max(Q)/q \leq \frac{p^{-1}n^{4/3}(\log n)^{C/3}}{q} = \frac{n^{4/3}(\log n)^{C/3}}{a}.$$

Thus

$$|I_z| \geq \frac{1}{4}\frac{(q_2L_2 - q_1L_1/4)a^{-1}}{\sqrt{\frac{q_2L_2+q_1L_1/8+t}{a}}}$$

$$|I_z| = \Omega(\frac{q_2L_2}{n^{2/3}(\log n)^{C/6}}). \tag{2.30}$$

By the definitions of I_x and I_z, we have, for any $x \in I_x$ and $z \in I_z$

$$0 \le az^2 + bz - q_1x - t \le a(z+1)^2 - q_1x - t \le q_2L_2. \tag{2.31}$$

Thus, for any such pair of x and z, if $az^2 + bz - q_1x - t$ is divisible by q_2, then $y := (az^2 + bz - q_1x - t)/q_2$ is an integer in $[1, L_2]$. We are now using the ideas from Section 2.7, with respect to modulo q_2 and the intervals I_x, I_z.

Let $\bar{q_1}$ be the reciprocal of q_1 modulo q_2 (recall that $(q_1, q_2) = 1$). Let f be a function given by Lemma 2.4.1 with respect to the interval I_x. For a given $z \in I_z$, the number of $x \in I_x$ satisfying (2.29) is at least N_z, where

$$N_z := \sum_{m \in \mathbf{Z}} f(\bar{q_1}az^2 + \bar{q_1}bz - \bar{q_1}t + mq_2).$$

By applying Poisson summation formula (2.8) and summing over z in I_z we obtain

$$N := \sum_{z \in I_z} N_z = \sum_{m \in \mathbf{Z}} \frac{1}{q_2} \widehat{f}(\frac{m}{q_2}) \sum_{z \in I_z} e(\frac{(\bar{q_1}az^2 + \bar{q_1}bz - \bar{q_1}t)m}{q_2}).$$

It suffices to show that $N > 0$. Similar to the proof of Lemma 2.4.3, we will again show that the right hand side is dominated by the contribution at $m = 0$. By triangle inequality, we have

$$|N - \frac{1}{q_2}\widehat{f}(0)|I_z|| \le \sum_{\substack{m \in \mathbf{Z} \\ m \ne 0}} \frac{1}{q_2} |\widehat{f}(\frac{m}{q_2})| |\sum_{z \in I_z} e(\frac{(\bar{q_1}az^2 + \bar{q_1}bz - \bar{q_1}t)m}{q_2})|.$$

Let γ be a sufficiently large constant and let

$$L' := \frac{8q_2(\log q_2)^\gamma}{L_1}.$$

We have

$$|N - \frac{1}{q_2}\widehat{f}(0)|I_z|| \le S_1 + S_2$$

where

$$S_1 := \sum_{|m| \ge L'} \frac{1}{q_2} |\widehat{f}(\frac{m}{q_2})| |\sum_{z \in I_z} e(\frac{(\bar{q_1}az^2 + \bar{q_1}bz - \bar{q_1}t)m}{q_2})|$$

and

$$S_2 := \sum_{\substack{|m|\leq L' \\ m\neq 0}} \frac{1}{q_2}|\widehat{f}(\frac{m}{q_2})|| \sum_{z\in I_z} e(\frac{(\bar{q}_1 az^2 + \bar{q}_1 bz - \bar{q}_1 t)m}{q_2})|.$$

To conclude the proof, we will show that both S_1 and S_2 are $o(\frac{\widehat{f}(0)|I_z|}{q_2})$.

Estimate for S_1. By the property of f,

$$S_1 \leq \frac{\widehat{f}(0)|I_z|}{q_2} \sum_{|m|\geq \frac{8q_2(\log q_2)^\gamma}{L_1}} \exp(-\delta\sqrt{|mL_1/(8q_2)|}).$$

By (2.17), and as q_2 is large ($q_2 \geq L_1 > n^{1/3}$), the inner sum is $o(1)$, so

$$S_1 = o(\frac{\widehat{f}(0)|I_z|}{q_2}) \qquad (2.32)$$

as desired.

Estimate for S_2. Let $q' = (\bar{q}_1 a, q_2)$. We can write

$$\bar{q}_1 a = q'q_1', q_2 = q'q_2' \text{ with } (q_1', q_2') = 1. \qquad (2.33)$$

Then

$$S_2 \leq \frac{\widehat{f}(0)}{q_2} \sum_{\substack{|m|\leq L' \\ m\neq 0}} |\sum_{z\in I_z} e(\frac{q_1' mz^2}{q_2'} + \frac{(\bar{q}_1 bz - \bar{q}_1 t)m}{q_2})|.$$

By Lemma 2.4.2 there are absolute constants c, α such that

$$S_2 \leq c\frac{\widehat{f}(0)}{q_2}\Big(L'\sqrt{|I_z|}(\log n)^\alpha + \frac{L'|I_z|(\log n)^\alpha}{\sqrt{q_2'}}\Big).$$

To show that $S_2 = o(\frac{\widehat{f}(0)|I_z|}{q_2})$, it suffices to show that

$$L'(\log n)^\alpha = o(\sqrt{|I_z|}) \qquad (2.34)$$

and

$$L'(\log n)^{\alpha} = o(q_2') \tag{2.35}$$

To verify (3.3), notice that by (2.30), we have

$$|I_z|L_1^2 = \Omega(\frac{L_1^2 q_2 L_2}{n^{2/3}(\log n)^{C/6}}).$$

Thus

$$\frac{|I_z|}{L'^2(\log n)^{2\alpha}} = \Omega(\frac{|I_z|L_1^2}{q_2^2(\log n)^{2\alpha+2\gamma}}) = \Omega\Big(\frac{L_1^2 L_2^2}{L_2 q_2 n^{2/3}(\log n)^{C/6+2\alpha+2\gamma}}\Big).$$

Since $(L_1L_2)^2 \geq (n(\log n)^{C/2})^2 = n^2 \log^C n$ and $L_2q_2 = O(\max(Q)) = O(p^{-1}n^{4/3}(\log n)^{C/3})$, the last formula is $\omega(1)$ if we set C sufficiently large compared to α and γ. This proves (3.3).

As a result,

$$\frac{\widehat{f}(0)}{q_2} L' \sqrt{|I_z|}(\log n)^{\alpha} = o(\widehat{f}(0)|I_z|/q_2).$$

Now we turn to (3.4). Recall that $q_2 = q'q_2'$ and $q' = (\bar{q_1}a, q_2) = (a, q_2)$ (as q_1 and q_2 are co-primes). Thus

$$q_2' \geq \frac{q_2}{a} = \frac{q_2}{pq}.$$

To show (3.4), it suffices to show that

$$\frac{q_2}{pq} = \omega(L'^2(\log n)^{2\alpha})$$

which (taking into account the definition of L') is equivalent to

$$q_2 L_1^2 = \omega(pqq_2^2(\log n)^{2\alpha+2\gamma}).$$

Multiplying both sides with $L_2q_2^{-1}$, it reduces to

$$L_1^2 L_2 = \omega(pqq_2L_2(\log n)^{2\alpha+2\gamma}).$$

Now we use the fact that $qq_2L_2 = O(\max(Q)) = O(p^{-1}n^{4/3}(\log n)^{C/3})$ and the lower bounds $L_1L_2 \geq n(\log n)^{C/2}$ and $L_1 \geq n^{1/3}(\log n)^{C/4}$. The claim follows by setting C sufficiently large compared to α and γ, as usual. Our proof is completed. $\square$

Chapter 3

The inverse Littlewood-Offord problem

3.1 Introduction

3.1.1 The Forward Littlewood-Offord problem

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

$$\rho(V) := \sup_x \mathbf{P}(S = x).$$

Motivated by their study of random polynomials, in the 1940s Littlewood and Offord [19] raised the question of bounding $\rho(V)$. (We call this the *forward* Littlewood-Offord problem, in contrast with the *inverse* Littlewood-Offord problem discussed in the next section.) They showed that $\rho(V) = O(n^{-1/2} \log n)$. Shortly after Littlewood-Offord paper, Erdős [6] gave a beautiful combinatorial proof of the refinement

$$\rho(V) \le \frac{\binom{n}{n/2}}{2^n} = O(n^{-1/2}). \tag{3.1}$$

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

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. Several results in this direction can be found in [13, 17].

Another direction was motivated by the observation that (3.1) can be improved significantly under additional assumptions on V. The first such result was discovered by Erdős and Moser [7], who showed that if v_i are distinct, then $\rho(V) = O(n^{-3/2}\log n)$. They conjectured that the logarithmic term is not necessary and this was confirmed by Sárközy and Szemerédi [27].

Theorem 3.1.2. *Let V be a set of n different integers, then*

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

In [15], Halász proved very general theorems that imply Theorem 3.1.2 and many others. One of his results can be formulated as follows.

Theorem 3.1.3. *Let l be a fixed integer and R_l be the number of solutions of the equation $v_{i_1} + \cdots + v_{i_l} = v_{j_1} + \cdots + v_{j_l}$. Then*

$$\rho(V) = O(n^{-2l-\frac{1}{2}} R_l).$$

It is easy to see, by setting $l = 1$, that Theorem 3.1.3 implies Theorem 3.1.2.

Another famous result in this area is that of Stanley [28], which, solving a conjecture of Erdős and Moser, shows when $\rho(V)$ attains its maximum under the assumption that the v_i are different.

Theorem 3.1.4. *Let n be odd and $V_0 := \{-\lfloor n/2\rfloor, \dots, \lfloor n/2\rfloor\}$. Then*

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

A similar result holds for the case n is even [28]. Stanley's proof of Theorem 3.1.4 used sophisticated machineries from algebraic geometry, in particular the hard-Lepschetz theorem. Few years later, a more elementary proof was given by Proctor [22]. This proof is also of algebraic nature, involving the representation of the Lie algebra $sl(2, \mathbf{C})$. As far as we know, there is no purely combinatorial proof.

It is natural to ask for the actual value of $\rho(V_0)$. From Theorem 3.1.2, one would guess (under the assumption that the elements of V are different) that

$$\rho(V_0) = (C_0 + o(1))n^{-3/2}$$

for some constant $C_0 > 0$. However, the algebraic proofs do not seem to yield the value of C_0. In fact, it is not obvious that $\lim_{n\to\infty} n^{3/2}\rho(V_0)$ exists.

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_0' has $\rho(V_0)'$ close to $C_0n^{-3/2}$, then should V_0', in some sense, close to V_0 ? (Notice that ρ is invariant under dilation.)

So far we have discussed various results on the concentration probability. There is another quantity that has been widely studied in theoretical probability, the small ball probability. This can be seen as the continuous analogue of the above. Instead of considering the probability that the random sum S concentrates on a single point, we consider the probability that it concentrates in a small ball.

Let $V = \{v_1, \dots, v_n\}$ be a multiset of n vectors in $\mathbf{R}^d$. For any $x \in \mathbf{R}^d$ and $r > 0$, we let $B(x, r)$ denote the closed disk of radius r centered at x. Let z be a real-valued random variable and $z_1, \dots, z_n$ be iid copies of z. We define the *small ball probability* as

$$\rho_{r,z}(V) := \sup_{x\in\mathbf{R}^d} \mathbf{P}(\sum_{i=1}^n v_i z_i \in B(x, r)).$$

Notice that, in contrast with the discrete setting, the small ball probability does not vary much if one slightly changes the vectors v_i.

The original setting in Littlewood-Offord paper [19] considered the bound of $\rho_{1,z}(V)$ when v_i are real numbers of absolute value at least 1, and z has Bernoulli distribution η. The continuous version of Erdős' theorem shows that in this case

$$\rho_{1,\eta}(V) \leq \frac{\binom{n}{n/2}}{2^n} = O(n^{-1/2}). \tag{3.2}$$

The results of Kleitman are also valid for this setting. Another beautiful extension given by Frankl and Füredi [8] also demonstrated a sharp upper bound for the small ball probability in general Euclidean space.

Theorem 3.1.5. *Assume that $v_1, \dots, v_n \in \mathbf{R}^d$ such that $\|v_i\|_2 \geq 1$. Then*

$$\rho_{r,\eta}(V) \leq (\lfloor r/2 \rfloor + 1 + o_d(1)) \frac{\binom{n}{n/2}}{2^n}.$$

3.1.6 The inverse Littlewood-Offord problem

We first discuss the discrete setting. Motivated by inverse theorems from additive combinatorics (see [40, Chapter 5]) and a variant for random sums in [33, Theorem 5.2], Tao and Vu [37] brought a different view to the problem. Instead of trying to improve the bound further by imposing new assumptions as done in the forward problems, they tried to provide the full picture by finding the underlying reason for the concentration probability to be large (say, polynomial in n).

Notice that the (multi)-set V has 2^n subsums, and $\rho(V) \geq n^{-C}$ mean that at least $\frac{2^n}{n^C}$ among these take the same value. This suggests that the set should have a very strong additive structure. In order to determine this structure, we first discuss a few examples of V where $\rho(V)$ is large.

Example 3.1.7. *Let $I = [-N, N]$ and $v_1, \dots, v_n$ be elements of I. Since $S \in nI$, by the pigeonhole principle, $\rho(V) \geq \frac{1}{|nI|} = \Omega(\frac{1}{nN})$. In fact, a short consideration yields a better bound. Notice that with probability at least .99, we have $S \in 10\sqrt{n}I$, thus again by the pigeonhole principle, we have $\rho(V) = \Omega(\frac{1}{\sqrt{n}N})$. If we set $N = n^{C-1/2}$ for some constant $C \geq 1/2$, then*

$$\rho(V) = \Omega(\frac{1}{n^C}). \tag{3.3}$$

The next, and more general, construction comes from additive combinatorics.

Example 3.1.8. *Let Q be a proper symmetric GAP of rank r and volume N. Let $v_1, \dots, v_n$ be (not necessarily distinct) elements of P. The random variable $S = \sum_{i=1}^n v_i \eta_i$ takes values in the GAP nP. Since $|nP| \leq \mathrm{Vol}(nB) = n^r N$, the pigeonhole principle implies that $\rho(V) \geq \Omega(\frac{1}{n^r N})$. In fact, using the same idea as in the previous example, one can improve the bound to $\Omega(\frac{1}{n^{r/2} N})$. If we set $N = n^{C-r/2}$ for some constant $C \geq r/2$, then*

$$\rho(V) = \Omega(\frac{1}{n^C}). \tag{3.4}$$

The above examples show that if the elements of V belong to a proper GAP with small rank and small cardinality then $\rho(V)$ is large. A few years ago, Tao and Vu [37] showed that this is essentially the only reason:

Theorem 3.1.9 (Weak inverse theorem,[37])**.** *Let $C, \epsilon > 0$ be arbitrary constants. There are constants r and C' depending on C and ϵ such that the following holds. Assume that $V = \{v_1, \dots, v_n\}$ is a multiset of integers satisfying $\rho(V) \geq n^{-C}$. Then there is a proper symmetric GAP Q of rank at most r and volume at most $n^{C'}$ which contains all but at most $n^{1-\epsilon}$ elements of V (counting multiplicity).*

Remark 3.1.10. *The presence of the small set of exceptional elements is not completely avoidable. For instance, one can add $o(\log n)$ completely arbitrary elements to V and only decrease $\rho(V)$ by a factor of $n^{-o(1)}$ at worst. Nonetheless we expect the number of such elements to be less than what is given by the results here.*

The reason we call Theorem 3.1.9 *weak* is that the dependence between the parameters is not optimal. In particular, they are far from reflecting the relations in (3.3) and (3.4). In a later paper [38], Tao and Vu refined the approach to obtain the following stronger result.

Theorem 3.1.11 (Strong inverse theorem, [38])**.** *Let C and $1 > \varepsilon$ be positive constants. Assume that*

$$\rho(V) \geq n^{-C}.$$

Then there exists a proper symmetric GAP Q of rank $d = O_{C,\varepsilon}(1)$ which contains all but $O_r(n^{1-\varepsilon})$ elements of V (counting multiplicity), where

$$|Q| = O_{C,\varepsilon}(n^{C-\frac{r}{2}+\varepsilon}).$$

The bound on $|Q|$ matches (3.4), up to the n^ϵ term. However, this error term seems to be the limit of the approach. The proofs of Theorem 3.1.9 and 3.1.11 rely on a replacement argument and various lemmas about random walks and GAPs.

Let us now consider another application of Theorem 3.1.11. Notice that Theorem 3.1.11 enables us to make very precise counting arguments. Assume that we would like to count the number of (multi)-sets V of integers with $\max|v_i| \leq N = n^{O(1)}$ such that $\rho(V) \geq \rho := n^{-C}$.

Fix $r \geq 1$, fix [1] a GAP Q with rank r and volume $|Q| = n^{C-\frac{r}{2}}$. The dominating term in the calculation will be the number of multi-subsets of size n of Q, which is

$$|Q|^n = n^{(C-\frac{r}{2}+\epsilon)n} \leq n^{Cn}n^{-\frac{n}{2}+\epsilon n} = \rho^{-n}n^{-n(\frac{1}{2}-\epsilon)}. \tag{3.5}$$

Motivated by questions from random matrix theory, Tao and Vu obtained the following continuous analogue of this result.

Let n be a positive integer and β, p be positive numbers that may depend on n. Let $\mathcal{S}_{n,\beta,\rho}$ be the collection of all multiple sets $V = (v_1, \ldots, v_n), v_i \in \mathbf{R}^2$ such that $\sum_{i=1}^n \|v_i\|^2 = 1$ and $\rho_{\beta,\eta}(V) \geq \rho$.

Theorem 3.1.12 (The β-net Theorem, [35]). *Let $0 < \epsilon \leq 1$ and $C > 0$ be constants. Then, for all sufficiently large n and $\beta \geq \exp(-n^{\epsilon/3})$ and $\rho \geq n^{-C}$ there is a set $\mathcal{S} \subset (\mathbf{R}^2)^n$ of size at most*

$$\rho^{-n}n^{-n(\frac{1}{2}-\epsilon)} + \exp(o(n))$$

such that for any $V = \{v_1, \ldots, v_n\} \in \mathcal{S}_{n,\beta,p}$ there is $V' = (v'_1, \ldots, v'_n) \in \mathcal{S}$ such that $\|v_i - v'_i\|_2 \leq \beta$ for all i.

The theorem looks a bit cleaner if we use $\mathbf{C}$ instead of $\mathbf{R}^2$ (as in [35]). However, we prefer the current form as it is more suitable for generalization. The set $\mathcal{S}$ is usually referred to as a β-net of $\mathcal{S}_{n,\beta,p}$.

Theorem 3.1.12 is at the heart of the establishment of the Circular Law conjecture in random matrix theory (see [35], also [36]). It also plays an important role in the study of condition number of randomly perturbed matrices (see [39]). Its proof in [35] is quite technical and occupies the bulk part of that paper.

[1] A more detailed version of Theorems 3.1.9 and 3.1.11 tells us that there are not too many ways to choose the generators of Q. In particular, if $N = n^{O(1)}$, the number of ways to fix these is negligible compared to the main term.

On the other hand, given the above discussion, one would, obviously, expects to obtain Theorem 3.1.12 as a simple corollary of a continuous analogue of Theorem 3.1.11. However, the arguments in [35] have not yet provided such inverse theorem (although it did provide a sufficient amount of information about the set S that makes the estimate possible). The paper [24] of Rudelson and Vershynin also contained a characterization of the set S, but their characterization of somewhat different spirit than those discussed in this chapter.

3.2 A new approach and new results

In this chapter, we present a new approach to the inverse theorem. The core of this new approach is a (long range) variant of Freiman's famous inverse theorem: Theorem 1.5.2.

The new approach seems useful. First of all, it enables us to remove the error term n^{ϵ} in Theorem 3.1.11, resulting in an optimal inverse theorem.

Theorem 3.2.1 (Optimal inverse Littlewood-Offord theorem, discrete case)**.** *Let C and $1 > \varepsilon$ be positive constants. There is a constant $c_1 = c_1(\varepsilon, C)$ such that the following holds. Assume that*

$$\rho(V) \geq n^{-C}.$$

Then there exists a proper symmetric GAP Q of rank $r = O_{C,\varepsilon}(1)$ which contains all but at most εn elements of V (counting multiplicity), where

$$|Q| = O_{C,\varepsilon}(\rho(V)^{-1} n^{-\frac{r}{2}}).$$

This immediately implies several forward theorems, such as Theorems 3.1.2 and 3.1.3.

Proof. (of Theorem 3.1.2) Assume, for contradiction, that there is a set V of n different number such that $\rho(V) \geq c_1 n^{-3/2}$ for some large constant c_1 to be chosen. Set $\varepsilon = .1, C = 3/2$. By Theorem 3.2.1, there is a GAP Q of rank r and size $O_\epsilon(\frac{1}{c_1} n^{C-\frac{r}{2}})$ that contains at least $.9n$ elements from V. This implies $|Q| \geq .9n$. By setting c_1 sufficiently large and using the fact that $C = 3/2$ and $r \geq 1$, we can guarantee that $|Q| \leq .8n$, a contradiction. □

Theorem 3.1.3 can be proved similarly with the detail left as an exercise.

Similar to [37, 38], our method and results can be extended (rather automatically) to much more general settings.

General V. Instead of taking V to be a subset of $\mathbf{Z}$, we can take it to be a subset of any abelian torsion free group G (thanks to Freiman isomorphism). We can also replace $\mathbf{Z}$ by the finite field $\mathbf{F}_p$, where p is any sufficiently large prime. (In fact, the first step in our proof is to embed V into $\mathbf{F}_p$.)

General η. We can replace the Bernoulli random variables by independent random variables η_i satisfying the following condition. There is a constant $c > 0$ and an infinite sequence of primes p such that for any p in the sequence, any (multi)-subset V of size n of $\mathbf{F}_p$ and any $t \in \mathbf{F}_p$

$$\prod_{i=1}^{n} |\mathbf{E}e_p(\eta_i v_i t)| \leq \exp(-c\sum_{i=1}^{n} \|\frac{v_i t}{p}\|^2) \tag{3.6}$$

where $\|x\|$ denote the distance from x to the closest integer (we view the elements of $\mathbf{F}_p$ as integers between 0 and $p-1$) .

Example 3.2.2. *(Lazy random walks) Given a parameter $0 < \mu \leq 1$, let η_i^μ be iid copies of a random variable η^μ: $\eta^\mu = 1$ or -1 with probability $\mu/2$, and $\eta^\mu = 0$ with probability $1-\mu$. The sum*

$$S^\mu(V) := \sum_{i=1}^{n} {\eta_i}^\mu v_i,$$

can be viewed as a lazy random walk with steps in V. A simple calculation shows

$$\mathbf{E}e_p(\eta x) = (1-\mu) + \mu \cos\frac{2\pi x}{p}.$$

It is easy to show that there is a constant $c > 0$ depending on μ such that

$$|(1-\mu) + \mu\cos\frac{2\pi x}{p}| \leq \exp(-c\|\frac{x}{p}\|^2).$$

Example 3.2.3. *(μ-bounded variables) It suffices to assume that there is some constant $0 < \mu \leq 1$ such that for all i*

$$|\mathbf{E}e_p(\eta_i x)| \leq (1-\mu) + \mu \cos \frac{2\pi x}{p}. \tag{3.7}$$

Theorem 3.2.4. *The conclusion of Theorem 3.2.1 holds for the case when V is a multi-subset of an arbitrary torsion free abelian group G and $\eta_i, 1 \leq i \leq n$ are independent random variables satisfying* (3.6)*.*

In the next application, we address the issues concerning Theorem 3.1.4. First, we compute the maximum concentration probability

$$\rho(V_0) = (\sqrt{\frac{24}{\pi}} + o(1))n^{-3/2}. \tag{3.8}$$

Next, we obtain a stable version of Theorem 3.1.3, which shows that if $\rho(V)$ is close to $(\sqrt{24/\pi} + o(1))n^{-3/2}$, then V is close to V_0 .

Theorem 3.2.5 (Asymptotic and stable Stanley theorem)**.** *Let V be a set of n distinct elements of a torsion free group, then*

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

Furthermore, there is a positive constant ϵ_0 such that for any constant $0 < \epsilon \leq \epsilon_0$, there is a constant $0 < \epsilon' = \epsilon'(\epsilon)$ such that $\epsilon' \to 0$ as $\epsilon \to 0$ and the following holds. If $V \subset \mathbf{Z}$ and $\rho(V) \geq (\sqrt{24/\pi} - \epsilon)n^{-3/2}$, then there exists an integer k which divides all $v \in V$ and

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

As a byproduct, we obtain the first non-algebraic proof for the asymptotic version of Stanley theorem. More importantly, this result 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.* It is easy to see that if

$$\sum_{v \in V} v^2 \leq (1+\epsilon') \sum_{v \in V_0} v^2$$

then $|V \backslash V_0| \leq \epsilon^{''} n$, with $\epsilon^{''}$ tends to zero with ϵ'. Our theorem actually says more than this, as it also bounds the elements that do not belong to V_0.

We now turn to the continuous setting. Let z be a real valued random variable such that there exists a constant C_z so

$$\mathbf{P}(1 \le z_1 - z_2 \le C_z) \ge 1/2, \tag{3.9}$$

where z_1, z_2 are idd copies of z. We notice that Bernoulli random variables are clearly of this type. (Also, the interested reader may find (3.9) more general than the condition of κ-controlled second moment defined in [35] and the condition of bounded third moment in [24].) In the above C_z is not uniquely defined. In what follows, we will take the smallest value of C_z.

We say that a vector $v \in \mathbf{R}^d$ is *δ-close* to a set $Q \subset \mathbf{R}^d$ if there exists a vector $q \in Q$ such that $\|v - q\|_2 \le \delta$. A set X is δ-close to a set Q if every element of X is δ-close to Q.

Theorem 3.2.6 (Continuous Inverse Littlewood-Offord theorem)**.** *Let $0 < \epsilon \le 1, 0 < C$ be constants. Let $\beta > 0$ be a parameter that may depend on n. Suppose that $V = \{v_1, \dots, v_n\}$ is a (multi-) subset of $\mathbf{R}^d$ such that $\sum_{i=1}^n \|v_i\|_2^2 = 1$ and that V has large small ball probability*

$$\rho := \rho_{\beta,z}(V) \ge n^{-C},$$

where z is a real random variable satisfying (3.9). Then the following holds. For any number n' between n^ϵ and n, there exists a proper symmetric GAP $Q = \{\sum_{i=1}^r x_i g_i : |x_i| \le L_i\}$ such that

- *(Full dimension) There exists $\sqrt{\frac{n'}{\log n}} \ll k' \ll \sqrt{n'}$ such that the dilate $P := \beta^{-1}k' \cdot Q$ contains the discrete hypercube $\{0,1\}^d$.*
- *(Approximation) At least $n - n'$ elements of V are $O(\frac{\beta}{k'})$-close to Q.*
- *(Small rank and cardinality) Q has constant rank $d \le r = O(1)$, and cardinality*

$$|Q| \le \max\left(O(\rho^{-1} n'^{(-r+d)/2}), 1\right).$$

- *(Small generators) There is a non-zero integer $p = O(\sqrt{n'})$ such that all steps g_i of Q have the form $g_i = (g_{i1}, \dots, g_{id})$, where $g_{ij} = \beta \frac{p_{ij}}{p}$ with $p_{ij} \in \mathbf{Z}$ and $p_{ij} = O(\beta^{-1}\sqrt{n'})$*

This theorem is sharp in the sense that the exponent $(-r+d)/2$ in the bound on $|Q|$ cannot be improved in general (see 3.5 for more details).

Theorem 3.2.6 implies the following corollary (see also 3.5 for a simple proof), from which one can derive Theorem 3.1.12 in a straightforward manner (similar to the discrete case discussed earlier).

Corollary 3.2.7. *Let $0 < \epsilon \le 1, 0 < C$ be constants. Let $\beta > 0$ be a parameter that may depend on n. Suppose that $V = \{v_1, \dots, v_n\}$ is a (multi-) subset of $\mathbf{R}^d$ such that $\sum_{i=1}^n \|v_i\|_2^2 = 1$ and that V has large small ball probability*

$$\rho := \rho_{\beta,z}(V) \ge n^{-C},$$

where z is a real random variable satisfying (3.9). Then the following holds. For any number n' between n^ϵ and n, there exists a proper symmetric GAP $Q = \{\sum_{i=1}^r x_i g_i : |x_i| \le L_i\}$ such that

- *At least $n - n'$ elements of V are β-close to Q.*

- *Q has small rank, $r = O(1)$, and small cardinality*

$$|Q| \le \max\left(O(\frac{\rho^{-1}}{\sqrt{n'}}), 1\right).$$

- *There is a non-zero integer $p = O(\sqrt{n'})$ such that all steps g_i of Q have the form $g_i = (g_{i1}, \dots, g_{id})$, where $g_{ij} = \beta \frac{p_{ij}}{p}$ with $p_{ij} \in \mathbf{Z}$ and $p_{ij} = O(\beta^{-1}\sqrt{n'})$*

In the above theorems, the hidden constants could depend on previously set constants ϵ, C, C_z, d. We could have written $O_{\epsilon,C,C_z,d}$ and $\ll_{\epsilon,C,C_z,d}$ everywhere, but these notations are somewhat cumbersome and this dependence is not our focus.

Proof. (of Theorem 3.1.12) Set $n' := n^{1-\frac{\epsilon}{2}}$. Let $\mathcal{S}'$ be the collection of all subsets of size at least $n - n'$ of GAPs whose parameters satisfy the conclusion of Theorem 3.2.6.

Since each GAP is determined by its generators and dimensions, the number of such GAPs is bounded by $((\beta^{-1}\sqrt{n'})\sqrt{n'})^{O(1)}(\frac{\rho^{-1}}{\sqrt{n'}})^{O(1)} = \exp(o(n))$. (The term $(\frac{\rho^{-1}}{\sqrt{n'}})^{O(1)}$ bounds the number of choices of the dimensions L_i.) Thus the cardinality of $\mathcal{S}'$ is at most $\Big((O_{d,C_z}(\frac{\rho^{-1}}{\sqrt{n'}}))^n + 1\Big)\exp(o(n))$.

We approximate each of the exceptional elements by a lattice point in $\beta \cdot (\mathbf{Z}/d)^d$. Thus if we let $\mathcal{S}''$ to be the set of these approximated tuples then $|\mathcal{S}''| \le \sum_{i\le n'}(O_d(\beta^{-1}))^i = \exp(o(n))$ (here we used the assumption that $\beta \ge \exp(-n^{\epsilon/3})$).

Set $\mathcal{S} := \mathcal{S}'\times\mathcal{S}''$. It is easy to see that $|\mathcal{S}| \le (n^{-1/2+\epsilon}\rho^{-1})^n + \exp(o(n))$. Furthermore, if $\rho(V) \ge n^{-O(1)}$ then V is β-close to an element of $\mathcal{S}$, concluding the proof. □

3.3 Proof of Theorem 3.2.1

Embedding. 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 $p \ge 2^n(\sum_{i=1}^n |v_i| + 1)$ suffices). If V is a subset of a general torsion-free group G, one can use Theorem 1.2.3.

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 $a \in \mathbf{F}_p$. We have

$$\rho = \mathbf{P}(S = a) = \mathbf{E}\frac{1}{p}\sum_{\xi\in\mathbf{F}_p} e_p(\xi(S-a)) = \mathbf{E}\frac{1}{p}\sum_{\xi\in\mathbf{F}_p} e_p(\xi S)e_p(-\xi a). \tag{3.10}$$

By independence

$$\mathbf{E}e_p(\xi S) = \prod_{i=1}^n e_p(\xi\eta_i v_i) = \prod_{i=1}^n \cos\frac{2\pi\xi v_i}{p} \tag{3.11}$$

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

$$\rho \leq \frac{1}{p}\sum_{\xi \in \mathbf{F}_p} |\cos \frac{2\pi v_i \xi}{p}| = \frac{1}{p}\sum_{\xi \in \mathbf{F}_p} |\frac{\cos \pi v_i \xi}{p}|. \tag{3.12}$$

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,

$$|\cos \frac{\pi x}{p}| \leq 1 - \frac{1}{2}\sin^2 \frac{\pi x}{p} \leq 1 - 2\|\frac{x}{p}\|^2 \leq \exp(-2\|\frac{x}{p}\|^2), \tag{3.13}$$

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

$$\rho \leq \frac{1}{p}\sum_{\xi \in \mathbf{F}_p}\prod_i |\cos \frac{\pi v_i \xi}{p}| \leq \frac{1}{p}\sum_{\xi \in F_p} \exp(-2\sum_{i=1}^{n} \|\frac{v_i \xi}{p}\|^2). \tag{3.14}$$

Large level sets. Now we consider the level sets $S_m := \{\xi | \sum_{i=1}^{n} \|v_i\xi/p\|^2 \leq m\}$. We have

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

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

$$|S_m|\exp(-m+2) \geq \rho p. \tag{3.15}$$

In fact, since $\rho \geq n^{-C}$, we can assume that $m = O(\log n)$.

Double counting and the triangle inequality. By double counting we have

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

So, for most v_i

$$\sum_{\xi \in S_m}\|\frac{v_i\xi}{p}\|^2 \leq \frac{C_0 m}{n}|S_m| \tag{3.16}$$

for some large constant C_0.

Set $C_0 = \varepsilon^{-1}$. By averaging, the set of v_i satisfying (3.16) has size at least $(1-\varepsilon)n$. We call this set V'. The set $V \backslash V'$ has size at most εn and this is the exceptional set that appears in Theorem 3.2.1. In the rest of the proof, we are going to show that V' is a dense subset of a proper GAP.

Since $\| \cdot \|$ is a norm, by the triangle inequality, we have for any $a \in kV'$

$$\sum_{\xi \in S_m} \| \frac{a\xi}{p} \|^2 \leq k^2 \frac{C_0 m}{n} |S_m|. \tag{3.17}$$

More generally, for any $l \leq k$ and $a \in lV'$

$$\sum_{\xi \in S_m} \| \frac{a\xi}{p} \|^2 \leq k^2 \frac{C_0 m}{n} |S_m|. \tag{3.18}$$

Dual sets. Define $S_m^* := \{a | \sum_{\xi \in S_m} \| \frac{a\xi}{p} \|^2 \leq \frac{1}{200} |S_m| \}$ (the constant 200 is adhoc and any sufficiently large constant would do). S_m^* can be viewed as some sort of a *dual* set of S_m. In fact, one can show as far as cardinality is concerned, it does behave like a dual

$$|S_m^*| \leq \frac{8p}{|S_m|}. \tag{3.19}$$

To see this, define $T(a) = \sum_{\xi \in S_m} \cos \frac{2\pi a \xi}{p}$. Using the fact that $\cos 2\pi z \geq 1 - 100 \|z\|^2$ for any $z \in \mathbf{R}$, we have, for any $a \in S_m^*$

$$T_a \geq \sum_{\xi \in S_m} (1 - 100 \| \frac{a\xi}{p} \|^2) \geq \frac{1}{2} |S_m|.$$

One the other hand, using the basic identity $\sum_{a \in \mathbf{F}_p} \cos \frac{2\pi a x}{p} = p \mathbf{I}_{x=0}$, we have

$$\sum_{a \in \mathbf{F}_p} T_a^2 \leq 2p |S_m|.$$

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

Set $k := c_1 \sqrt{\frac{n}{m}}$, for a properly chosen constant $c_1 = c_1(C_0)$. By (3.18) we have $\cup_{l=1}^k lV' \subset S_m^*$. Set $V'' = V' \cup \{0\}$; we have $kV'' \subset S_m^* \cup \{0\}$. This results in the critical bound

$$|kV^{''}| = O(\frac{p}{|S_m|}) = O(\rho^{-1}\exp(-m+2)). \tag{3.20}$$

The Long Range Inverse Theorem. The role of $\mathbf{F}_p$ is now no longer important, so we can view v_i as integers. The inequality (3.20) is exactly the assumption of our Long Range Inverse Theorem.

With this theorem in hand, we are ready to conclude the proof. A slight technical problem is that $V^{''}$ is not a set but a multi-set, so we are going to apply Theorem 1.5.2 to X being the set of different elements of $V^{''}$. Notice that $k = \Omega(\sqrt{\frac{n}{m}}) = \Omega(\sqrt{\frac{n}{\log n}})$, so $\rho^{-1} \le n^C$ is bounded from above by k^{2C+1}.

It follows from Theorem 1.5.2 that X is a subset of a proper symmetric GAP Q of rank $r = O_{C,\epsilon}(1)$ and cardinality

$$\begin{aligned} O_{C,\epsilon}(k^{-r}|kX|) = O_{C,\epsilon}(k^{-r}|kV^{''}|) &= O_{C,\epsilon}\left(\rho^{-1}\exp(-m)(\sqrt{\frac{n}{m}})^{-r}\right) \\ &= O_{C,\epsilon}(\rho^{-1}n^{-r}), \end{aligned}$$

concluding the proof.

3.4 Proof of Theorem 3.2.6

We denote the z-norm of a real number to be

$$\|w\|_z := (\mathbf{E}\|w(z_1 - z_2)\|^2)^{1/2},$$

where z_1, z_2 are two iid copies of z.

Fourier analysis. Our first step is to obtain the following analogue of (3.14), using the Fourier transform.

Lemma 3.4.1 (bounds for small ball probability)**.**

$$\rho_{r,z}(V) \le \exp(\pi r^2)\int_{\mathbf{R}^d}\exp(-\sum_{i=1}^{n}\|\langle v_i, \xi\rangle\|_z^2/2 - \pi\|\xi\|_2^2)d\xi.$$

This lemma is basically from [35]; the proof is presented here for the reader's convenience.

Proof. (of Lemma 3.4.1) We have

$$\begin{aligned}\mathbf{P}(\sum_{i=1}^n z_i v_i \in B(x,r)) &= \mathbf{P}(\|\sum_{i=1}^n z_i v_i - x\|_2^2 \le r^2)\\ &= \mathbf{P}\left(\exp(-\pi\|\sum_{i=1}^n z_i v_i - x\|_2^2) \ge \exp(-\pi r^2)\right)\\ &\le \exp(\pi r^2)\mathbf{E}\exp(-\pi\|\sum_{i=1}^n z_i v_i - x\|_2^2).\end{aligned}$$

Notice that

$$\exp(-\pi\|x\|_2^2) = \int_{\mathbf{R}^d} e(\langle x, \xi\rangle)\exp(-\pi\|\xi\|_2^2)d\xi.$$

We thus have

$$\mathbf{P}(\sum_{i=1}^n z_i v_i \in B(x,r)) \le \exp(\pi r^2)\int_{\mathbf{R}^d}\mathbf{E}e(\langle\sum_{i=1}^n z_i v_i, \xi\rangle)e(-\langle x, \xi\rangle)\exp(-\pi\|\xi\|_2^2)d\xi.$$

Using

$$|\mathbf{E}e(\langle\sum_{i=1}^n z_i v_i, \xi\rangle)| = \prod_{i=1}^n |\mathbf{E}e(z_i\langle v_i, \xi\rangle)|,$$

and

$$|\mathbf{E}e(z_i\langle v_i, \xi\rangle)| \le |\mathbf{E}e(z_i\langle v_i, \xi\rangle)|^2/2 + 1/2 \le \exp(-\|\langle v_i, \xi\rangle\|_z^2/2),$$

we obtain

$$\rho_{r,z}(V) \le \exp(\pi r^2)\int_{\mathbf{R}^d}\exp(-\sum_{i=1}^n \|\langle v_i, \xi\rangle\|_z^2/2 - \pi\|\xi\|_2^2)d\xi.$$

□

Next, consider $V_\beta := \beta^{-1}\cdot V = \{\beta^{-1}v_1, \ldots, \beta^{-1}v_n\}$. It is clear that

$$\rho_{\beta,z}(V) = \rho_{1,z}(V_\beta).$$

We now work with V_β. Thus $\rho_{1,z}(V_\beta) \geq n^{-O(1)}$ and $\sum_{v \in V_\beta} \|v\|^2 = \beta^{-2}$.

For short, we write ρ for $\rho_{1,z}(V_\beta)$. Set $M := 2A \log n$ where A is large enough. From Lemma 3.4.1 and that $\rho \geq n^{-O(1)}$ we easily obtain

$$\int_{\|\xi\|_2 \leq M} \exp(-\frac{1}{2}\sum_{v \in V_\beta} \|\langle v, \xi\rangle\|_z^2 - \pi\|\xi\|_2^2) d\xi \geq \frac{\rho}{2}.$$

Large level sets. For each integer $0 \leq m \leq M$ we define the level set

$$S_m := \left\{ \xi \in \mathbf{R}^d : \sum_{v \in V_\beta} \|\langle v, \xi\rangle\|_z^2 + \|\xi\|_2^2 \leq m \right\}.$$

Then it follows that $\sum_{m \leq M} \mu(S_m) \exp(-\frac{m}{2} + 1) \geq \rho$, where $\mu(.)$ denotes the Lebesgue measure of a measurable set. Hence there exists $m \leq M$ such that $\mu(S_m) \geq \rho \exp(\frac{m}{4} - 2)$.

Next, since $S_m \subset B(0, \sqrt{m})$, by pigeonhole principle there exists a ball $B(x, \frac{1}{2}) \subset B(0, \sqrt{m})$ such that

$$\mu(B(x, \frac{1}{2}) \cap S_m) \geq c_d \mu(S_m) m^{-d/2} \geq c_d \rho \exp(\frac{m}{4} - 2) m^{-d/2}.$$

Consider $\xi_1, \xi_2 \in B(x, 1/2) \cap S_m$. By Cauchy-Schwarz inequality, and notice that $\|.\|_z$ is a norm, we have

$$\sum_{v \in V_\beta} \|\langle v, (\xi_1 - \xi_2)\rangle\|_z^2 \leq 4m.$$

Since $\xi_1 - \xi_2 \in B(0,1)$ and $\mu(B(x, \frac{1}{2}) \cap S_m - B(x, \frac{1}{2}) \cap S_m) \geq \mu(B(x, \frac{1}{2}) \cap S_m)$, if we put

$$T := \{\xi \in B(0,1), \sum_{i=1}^n \|\langle \xi, v_i\rangle\|_z^2 \leq 4m\},$$

then

$$\mu(T) \geq c_d \rho \exp(\frac{m}{4} - 2) m^{-d/2}.$$

Discretization. Choose N to be a sufficiently large prime (depending on the set T). Define the discrete box

$$B_0 := \{(k_1/N, \dots, k_d/N) : k_i \in \mathbf{Z}, -N \le k_i \le N\}.$$

We consider all the shifted boxes $x + B_0$, where $x \in [0, 1/N]^d$. By pigeonhole principle, there exists x_0 such that the size of the discrete set $(x_0 + B_0) \cap T$ is at least the expectation, $|(x_0 + B_0) \cap T| \ge N^d \mu(T)$ (to see this, we first consider the case when T is a box itself).

Let us fix a $\xi_0 \in (x_0 + B_0) \cap T$. Then for any $\xi \in (x_0 + B_0) \cap T$ we have

$$\sum_{v \in V_\beta} \|\langle v, \xi_0 - \xi\rangle\|_z^2 \le 2\left(\sum_{v \in V_\beta} \|\langle v, \xi\rangle\|_z^2 + \sum_{v \in V_\beta} \|\langle v, \xi_0\rangle\|_z^2\right) \le 16m.$$

Notice that $\xi_0 - \xi \in B_1 := B_0 - B_0 = \{(k_1/N, \dots, k_d/N) : k_i \in \mathbf{Z}, -2N \le k_i \le 2N\}$. Thus there exists a subset S of size at least $c_d N^d \rho \exp(\frac{m}{4} - 2) m^{-d/2}$ of B_1 such that the following holds for any $s \in S$

$$\sum_{v \in V_\beta} \|\langle v, s\rangle\|_z^2 \le 16m.$$

Double counting. We let $y = z_1 - z_2$, where z_1, z_2 are iid copies of z. By definition of S, we have

$$\sum_{s \in S} \sum_{v \in V_\beta} \|\langle v, s\rangle\|_z^2 \le 16m|S|$$

$$\mathbf{E}_y \sum_{s \in S} \sum_{v \in V_\beta} \|y\langle v, s\rangle\|_{\mathbf{R}/\mathbf{Z}}^2 \le 16m|S|.$$

It is then implied that there exists $1 \le |y_0| \le C_z$ such that

$$\sum_{s \in S} \sum_{v \in V_\beta} \|y_0\langle v, s\rangle\|_{\mathbf{R}/\mathbf{Z}}^2 \le 16m|S|\mathbf{P}(1 \le |y| \le C_z)^{-1}.$$

On the other hand, by property (3.9) we have $\mathbf{P}(1 \le |y| \le C_z) \ge 1/2$. So

$$\sum_{s \in S} \sum_{v \in V_\beta} \|y_0\langle v, s\rangle\|_{\mathbf{R}/\mathbf{Z}}^2 \le 32m|S|.$$

Let n' be any number between n^ϵ and n. We say that $v \in V_\beta$ is *bad* if

$$\sum_{s\in S} \|y_0\langle v, s\rangle\|_{\mathbf{R}/\mathbf{Z}}^2 \geq \frac{32m|S|}{n'}.$$

Then the number of bad vectors is at most n'. Let V'_β be the set of remaining vectors. Thus V'_β contains at least $n - n'$ elements. In the rest of the proof, we are going to show that V'_β is close to a GAP, as claimed in the theorem.

Dual sets. Consider $v \in V'_\beta$, we have $\sum_{s\in S} \|y_0\langle s, v\rangle\|_{\mathbf{R}/\mathbf{Z}}^2 \leq 32|K|/n'$.

Set $k := \sqrt{\frac{n'}{64\pi^2 m}}$ and let $V''_\beta := k(V'_\beta \cup \{0\})$. By Cauchy-Schwarz inequality (see (3.18)), for any $a \in V''_\beta$ we have

$$\sum_{s\in S} 2\pi^2 \|\langle s, y_0 a\rangle\|_{\mathbf{R}/\mathbf{Z}}^2 \leq \frac{|S|}{2},$$

which implies

$$\sum_{s\in S} \cos(2\pi\langle s, y_0 a\rangle) \geq \frac{|S|}{2}.$$

Observe that for any $x \in B(0, \frac{1}{256d})$ and any $s \in S \subset B(0,2)$ we always have $\cos(2\pi\langle s, x\rangle) \geq 1/2$ and $\sin(2\pi\langle s, x\rangle) \leq 1/12$. Thus for any $x \in B(0, \frac{1}{256d})$,

$$\sum_{s\in S} \cos\left(2\pi\langle s, (y_0 a + x)\rangle\right) \geq \frac{|S|}{4} - \frac{|S|}{12} = \frac{|S|}{6}.$$

On the other hand,

$$\begin{aligned}\int_{x\in[0,N]^d} \left(\sum_{s\in S} \cos(2\pi\langle s, x\rangle)\right)^2 dx &\leq \sum_{s_1,s_2\in S} \int_{x\in[0,N]^d} \exp\left(2\pi\sqrt{-1}\langle s_1 - s_2, x\rangle\right) dx \\ &\ll_d |S|N^d.\end{aligned}$$

Hence we deduce the following

$$\mu_{x\in[0,N]^d}\left((\sum_{s\in S} \cos(2\pi\langle s, x\rangle))^2 \geq (\frac{|S|}{6})^2 \right) \ll_d \frac{|S|N^d}{(|S|/6)^2} \ll_d \frac{N^d}{|S|}.$$

Now we use the fact that S has large size, $|S| \gg_d N^d \rho \exp(\frac{m}{4} - 2)m^{-d/2}$, and $y_0 V''_\beta + B(0, \frac{1}{256d}) \subset [0,N]^d$,

$$\mu(y_0 V''_\beta + B(0, \frac{1}{256d})) \ll_d \rho^{-1} \exp(-\frac{m}{4} + 2)m^{d/2}.$$

Thus, we obtain the following analogue of (3.20)

$$\mu\left(k(V'_\beta \cup \{0\}) + B(0, \frac{1}{256dy_0})\right) \ll_d \rho^{-1} y_0^{-d} \exp(-\frac{m}{4} + 2)m^{d/2}. \tag{3.21}$$

The Long Range Inverse Theorem. Our analysis now again relies on the Long Range Inverse Theorem. Let $D := 1024dy_0$. We approximate each vector v' of V'_β by a closest vector in $(\frac{\mathbf{Z}}{Dk})^d$,

$$\|v' - \frac{a}{Dk}\|_2 \leq \frac{\sqrt{d}}{Dk}, \text{ with } a \in \mathbf{Z}^d.$$

Let A_β be the collection of all such a. Since $\sum_{v' \in V'_\beta} \|v'\|_2^2 = O(\beta^{-2})$, we have

$$\sum_{a \in A_\beta} \|a\|_2^2 = O_{d,C_z}(k^2 \beta^{-2}). \tag{3.22}$$

It follows from (3.21) that

$$\begin{aligned} |k(A_\beta + C(0,1))| &= O_{d,C_z}\left(\rho^{-1}(Dk)^d y_0^{-d} \exp(-\frac{m}{4} + 2)m^{d/2}\right) \\ &= O_{d,C_z}\left(\rho^{-1} k^d \exp(-\frac{m}{4} + 2)m^{d/2}\right), \end{aligned}$$

where $C(0,r)$ is the discrete cube $\{(z_1, \dots, z_d) \in \mathbf{Z}^d : |z_i| \leq r\}$.

Now we apply Theorem 1.5.2 to the set $A_\beta + C(0,1)$ (notice that $0 \in A_\beta$). There exists a proper GAP $P = \{\sum_{i=1}^r x_i g_i : |x_i| \leq L_i\} \subset \mathbf{Z}^d$ containing $A_\beta + C(0,1)$ which has small rank $r = O(1)$, and small size

$$\begin{aligned} |P| &= O_{d,C_z}\left((\rho^{-1} k^d \exp(-\frac{m}{4} + 2)m^{d/2} k^{-r}\right) \\ &= O_{d,C_z}(n'^{(d-r)/2} \rho^{-1}). \end{aligned}$$

Moreover, we have learned from the proof of Theorem 1.5.2 and Lemma 1.1.2 that kQ can be contained in a set $ck(A_\beta+C(0,1))$ for some $c=O(1)$. Using (3.22), we conclude that all the generators g_i of Q are bounded,

$$\|g_i\|_2 = O_{d,C_z}(k\beta^{-1}).$$

Next, since $C(0,1)\subset Q$, the rank r of P is at least d. It is a routine calculation to see that $Q := \frac{\beta}{Dk}\cdot P$ satisfies all required properties in the theorem.

3.5 Remarks on Theorem 3.2.6

Consider the set $U := [-2n,-n]\cup[n,2n]$. Sample n points $v_1,\dots,v_n$, from U, independently with respect to the (continuous) uniform distribution and let A be the set of sampled points. Let ξ be the gaussian random variable $N(0,1)$ and consider the sum

$$S := v_1\xi_1+\cdots+v_n\xi_n,$$

where ξ_i are iid copies of ξ.

S has gaussian distribution with mean 0 and variance $\Theta(n^3)$, with probability one. Thus, for any interval I of length 1, $\mathbf{P}(S\in I)\le Cn^{-3/2}$, for some constant C.

Set $n'=\delta n$, for some small positive constant δ. Theorem 3.2.6 states that (most of) A is $O(\frac{\log n}{\sqrt{n}})$-close to a GAP of rank r and volume $O(n^{2-\frac{r}{2}})$. We show that one cannot replace this bound by $O(n^{2-\frac{r}{2}-\epsilon})$. There are only three possible values for r: $r=1,2,3$ and our claim follows from the following simple lemma.

Lemma 3.5.1. *Let C,δ,ϵ be positive constants and $n\to\infty$. The followings hold with probability $1-o(1)$ (with respect to the random choice of A).*

- *There is no subset A' of A of cardinality at least $(1-\delta)n$ and an AP Q of length at most $Cn^{3/2-\epsilon}$ such that A' is $\frac{C\log n}{\sqrt{n}}$-close to Q.*
- *There is no subset A' of A of cardinality at least $(1-\delta)n$ and a GAP Q of rank 2 and volume at most $Cn^{1-\epsilon}$ such that A' is $\frac{C\log n}{\sqrt{n}}$-close to Q.*

- *There is no subset A' of A of cardinality at least $(1-\delta)n$ and a GAP Q of rank 3 and volume at most $Cn^{1/2-\epsilon}$ such that A' is $\frac{C\log n}{\sqrt{n}}$-close to Q.*

The above construction can be generalized to higher dimensions as well, but we do not attempt to do so here. In the rest of this section, we prove Corollary 3.2.7.

Proof. We consider the following two cases.

Case 1: $r \geq d+1$. Consider the GAP P at the end of the proof of Theorem 3.2.6. Recall that $|P| = O_{d,C_z}(\rho^{-1}n'^{(d-r)/2}) = O_{d,C_z}(\rho^{-1}/\sqrt{n'})$. Let

$$Q := \frac{\beta}{Dk} \cdot P.$$

It is clear that Q satisfies all the conditions of Corollary 3.2.7. (Notice that in this case we obtain a stronger approximation: almost all elements of V are $O(\frac{\beta \log n'}{\sqrt{n'}})$-close to Q.)

Case 2: $r = d$. Because the unit vectors $e_j = (0,\dots,1,\dots,0)$ belong to $P = \{\sum_{i=1}^{d} x_i g_i : |x_i| \leq N_i\} \subset \mathbf{Z}^d$, the set of generators $g_i, i = 1,\dots,d$ forms a base with unit determinant of $\mathbf{R}^d$. In P, consider the set of lattice points with all coordinates divisible by k. We observe that (for instance by [40, Theorem 3.36]) this set can be contained in a GAP P' of rank d and cardinality $O(\frac{1}{k^r}|P|) = O(\rho^{-1}/n'^{r/2})$ (here we use the bound $|P| = O(\rho^{-1}\exp(-\frac{m}{4})m^{d/2})$). Next, define

$$Q := \frac{\beta}{Dk} \cdot P'.$$

It is easy to verify that Q satisfies all the conditions of Corollary 3.2.7. (Notice that in this case we obtain a stronger bound on the size of Q.) □

3.6 Proof of the optimal asymptotic bound (3.8)

Consider V_0 as in Theorem 3.2.5 and view the elements of $\mathbf{F}_p$ as integers between $-p/2$ and $p/2$. By (3.10), we have

$$\mathbf{P}(S=0) = \frac{1}{p}\sum_{\xi\in\mathbf{F}_p}\prod_{i\in V_0}\cos\frac{2\pi i\xi}{p} = \sum_{\xi\in\mathbf{F}_p}\prod_{i\in V_0}\cos\frac{2\pi i\xi}{p}. \qquad (3.23)$$

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 3.6.1.

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

Lemma 3.6.2.

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

The two lemmas together imply that $\rho(V_0)\ge(\sqrt{\frac{24}{\pi}}+o(1))n^{-3/2}$. The matching lower bound also follows from these lemmas and (3.12). This verifies (3.8).

Proof. (of Lemma 3.6.1) 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 $n^{-3/2}p\log^2 n$, 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

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

completing the proof. □

Proof. (of Lemma 3.6.2). To prove Lemma 3.6.2 we use the upper bound from (3.13). We split the sum into three subsums, according to the magnitude of $\|\xi\|$. We make frequently use of the simple fact that if $|w|\|\frac{\xi}{p}\| \le 1/2$ then $\|\frac{w\xi}{p}\| = |w|\|\frac{\xi}{p}\|$.

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

$$\sum_{w\in V_0}\|\frac{w\xi}{p}\|^2 = \sum_{w\in V_0} w^2\|\frac{\xi}{p}\|^2 \ge \frac{n^3}{12}\frac{\log^4 n}{n^3} \ge 4\log n \tag{3.24}$$

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

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

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

For any ξ in this case, let $W(\xi)$ be the set of pairwise disjoint pairs $(w,w') \in V_0$ that maximizes the sum $M(\xi) := \sum_{(w,w')\in W(\xi)}|w-w'|^2$ under the constrain that $|w-w'|\|\frac{\xi}{p}\| \le \frac{1}{2}$ for all $(w,w') \in W(\xi)$. It is easy to check that $M(\xi) \ge n^c$ for some constant $c > 0$ for all ξ in this case (For more details see Lemma 3.9.2.) From here one can conclude that the contribution of this subsum is at most $\exp(n^{-\Omega(1)}) = o(n^{-4})$.

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

$$\sum_{w\in V_0}\|\frac{w\xi}{p}\|^2 \ge \frac{n}{8}, \tag{3.25}$$

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

□

3.7 Proof of Theorem 3.2.5

The proof of this theorem has two steps. In the first step, we refine our approach to prove the following theorem.

Theorem 3.7.1 (Characterization of sets near optimal concentration probability)**.** *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 $V = \{v_1, \dots, v_n\}$ be a subset of size n of $\mathbf{F}_p$, where $p \gg n$ is a large prime, such that $\rho(V) \geq \delta n^{-3/2}$. Then there exists a number $k \in \mathbf{F}_p$ and a partition $V = V_1 \cup V_2$ with the following properties:*

- $|V_1| \leq C_1$,
- $\sum_{v \in V_2} \| \frac{k^{-1}v}{p} \|^2 \leq \frac{C_2 n^3}{p^2}$.

This can be seen as a more precise version of Theorem 3.2.1 in the case $C = 3/2$. The appearance of k is necessary as $\rho(V)$ is invariant under dilation.

One can easily derive from this theorem a similar statement for a set V of integers.

Corollary 3.7.2. *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 $V = \{v_1, \dots, v_n\}$ be a subset of size n of $\mathbf{Z}$ such that $\rho(V) \geq \delta n^{-3/2}$. Then there exists a number $k \in \mathbf{Z}$ and a partition $V = V_1 \cup V_2$ with the following properties:*

- $|V_1| \leq C_1$,
- *k divides all elements of V_2, and $\sum_{v \in V_2} |\frac{v}{k}|^2 \leq C_2 n^3$.*

Let us now sketch the second step. Corollary 3.7.2 implies that most of the elements of V (after dividing by k) belong to the interval $[-C_3 n, C_3 n]$ for some large constant C_3. For the sake of discussion, assume that all elements of V are in this interval. Then we can finish the proof by applying the analysis in the previous section to V the same way we did with V_0. The fact that V now in $[-C_3 n, C_3 n]$ instead of $[-n/2, n/2]$ has little importance. As far as V has constant density, all arguments will extend with slight modifications. As a result, at the end we will have

$$\rho(V) \leq (1 + o(1))\Sigma_1,$$

where

$$\Sigma_1 = \frac{1}{p} \sum_{\|\frac{\xi}{p}\| \leq \frac{\log^2 n}{n^{3/2}}} \prod_{w \in V} \cos \frac{\pi w \xi}{p}.$$

Use the Taylor expansion and exponential approximation as in the previous section, the right hand side is

$$(1+o(1)) \int_{|x| \leq \frac{\log^2 n}{n^{3/2}}} \exp\left(-(1/2+o(1))\pi^2 \sum_{w \in V} w^2 x^2\right).$$

By the gaussian identity and change of variables, this is

$$(\sqrt{\frac{24}{\pi}} + o(1))n^{-3/2} \frac{n^3/12}{\sum_{w \in V} w^2}.$$

It follows that if $\rho(V) \geq (1-\epsilon)\sqrt{\frac{24}{\pi}} n^{-3/2}$ then $\frac{n^3/12}{\sum_{w \in V} w^2} \geq 1 - \frac{3}{2}\varepsilon$. This implies

$$\sum_{w \in V} w^2 \leq (1+2\varepsilon)\frac{n^3}{12},$$

giving the desired claim.

In what follows, we complete the above two steps in details. In the next section, we prove Theorem 3.7.1. In Section 3.9, we fill in the details of the second steps, which include the necessary modifications of the arguments from the previous section and the treatment of exceptional elements.

3.8 Proof of Theorem 3.7.1

We use a well-known result from Additive Combinatorics.

Theorem 3.8.1 (Cauchy-Davenport)**.** *[40, Theorem 5.4] Assume that $A, B \subset \mathbf{F}_p$. Then $|A+B| \geq \min(p, |A|+|B|-1)$.*

We also a result that allows us to pass from $\mathbf{F}_p$ back to torsion-free groups ([11, Theorem 1.3]).

Theorem 3.8.2. *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 $\mathbf{Z}$.*

Now we are ready to prove Theorem 3.7.1. By (3.15)

$$|S_m| \geq \exp(\frac{m}{4}-2)\rho p \geq \delta \exp(\frac{m}{4}-2)pn^{-3/2}. \tag{3.26}$$

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

Structure of S_m. Consider a set sequence, $S_m, 2S_m, \dots, 2^l S_m$, where l is the largest integer such that $4^l m \leq n/100$.

Assume that i_0 is the smallest index such that $|2^i S_m| \geq 2.1|2^{i-1}S_m|$ for all $1 \leq i \leq i_0$. Thus $|2^{i_0}S_m| \geq (2.1)^{i_0}|S_m|$.

By Cauchy-Schwartz inequality and by the definition of level sets, $kS_m \subset S_{k^2 m}$ holds for all k. In particular, $2^l S_m \subset S_{4^l m} \subset S_{n/100}$.

On the other hand, by Theorem 3.8.1, $|2^l S_m| \geq 2^{(l-i_0)}(|2^{i_0}S_m| - 1)$. Hence

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

Observe that $S_{n/100}$ is the dual set of V (see Section 3.3), so $|S_{n/100}| \leq 8p/n$. Insert this estimate and $2^l = \Theta(\sqrt{n/m})$ into the above inequalities, we derive that $m = O_\delta(1)$ and $i_0 = O_\delta(1)$.

Now we consider the set $X := 2^{i_0}S_m$. By definition, for $|X| \leq |S_{n/100}| \ll p/n$, Theorem 3.8.2 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 1.3.2, and then Lemma 1.1.2 (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 1 and of size $|P| = \Theta(|X|) \gg p/n^{3/2}$. Since X is symmetric, we may assume that P is symmetric.

Set $M := h^2 4^{i_0} m = O_\delta(1)$. Then $P \subset hX = h2^{i_0}S_m \subset S_{h^2 4^{i_0} m} = S_M$.

To summarize, we have showed that there exists a constant $C = O_\delta(1)$ and a symmetric arithmetic progression $P = \{kx : |x| \le \frac{Cp}{n^{3/2}}\}$ such that the following holds for all $\xi \in P$

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

Our next step is to derive structure for V. By dilating the elements of V by k^{-1}, we could assume that $\sum_{i=1}^{n} \|\frac{jv_i}{p}\|^2 \le M$ for all $0 \le j \le \frac{Cp}{n^{3/2}}$.

Summing over j we get

$$\sum_{i=1}^{n} \sum_{0 \le j \le \frac{Cp}{n^{3/2}}} \|\frac{jv_i}{p}\|^2 \le \frac{CMp}{n^{3/2}}. \tag{3.27}$$

Set $V_1 := \{v \in V : \|\frac{v}{p}\| \ge \frac{C'n^{3/2}}{p}\}$ for some sufficiently large constant C'. We next show that V_1 contains only a few elements.

Viewing $\mathbf{F}_p$ as $\{0, \dots, p-1\}$, we observe that if $v \in V_1$ then $C'n^{3/2} \le v \le p - C'n^{3/2}$. Provided that CC' is sufficient large, and j varies in $0 \le j \le \frac{Cp}{n^{3/2}}$, we conclude that there are at least $\frac{Cp}{3n^{3/2}}$ indices satisfying $jv \in [\frac{p}{4}, \frac{3p}{4}]$ (in $\mathbf{F}_p$). It follows that

$$\sum_{0 \le j \le \frac{Cp}{n^{3/2}}} \|\frac{jv}{p}\|^2 \ge (\frac{1}{4})^2 \frac{Cp}{3n^{3/2}}.$$

Summing over $v \in V_1$ we obtain

$$\sum_{v \in V_1} \sum_{0 \le j \le \frac{Cp}{n^{3/2}}} \|\frac{jv}{p}\|^2 \ge \frac{C}{48} \frac{|V_1| p}{n^{3/2}}.$$

Together with (3.27), this implies that $|V_1| \le 48M$.

Next we consider $V_2 := V \backslash V_1$. By definition, $\|\frac{jv}{p}\| = j\|\frac{v}{p}\|$ for all $j \le \frac{p}{2C'n^{3/2}}$ and all $v \in V_2$. Thus

$$\begin{aligned}\sum_{0 \le j \le \frac{p}{n^{3/2}}} \|\frac{jv}{p}\|^2 &\ge \sum_{0 \le j \le \frac{p}{2C'n^{3/2}}} j^2 \|\frac{v}{p}\|^2 \\ &\ge \frac{1}{64C'^3} \frac{p^3}{n^{9/2}} \|\frac{v}{p}\|^2.\end{aligned}$$

Summing over $v \in V_2$ and using (3.27), we obtain

$$\sum_{v\in V_2} \frac{1}{64C'^3}\frac{p^3}{n^{9/2}}\|\frac{v}{p}\|^2 \leq \sum_{v\in V_2}\sum_{0\leq j\leq \frac{Cp}{n^{3/2}}} \|\frac{jv}{p}\|^2$$
$$\leq \frac{CMp}{n^{3/2}}.$$

Hence, $\sum_{v\in V_2}\|v/p\|^2 \leq 64CC'^3Mn^3/p^2$, concluding the proof.

3.9 Completing the proof of Theorem 3.2.5: details of the second step

By applying Theorem 3.7.1, we obtain a partition $V = V_1 \cup V_2 = V_1 \cup k \cdot W_2$, where $W_2 = k^{-1}V_2$ and

- $|V_1| \leq C_1$,

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

Let $C = C(\epsilon)$ be a large positive constant and $c = c(\epsilon)$ be a small positive constant. By setting C, c properly and throwing away a small amount of elements of W_2, we can assume that W_2 has the following properties

- $|W_2| \geq (1-\epsilon)n$;

- $W_2 \subset [-Cn, Cn]$;

- W_2 is *c-irreducible*, i.e. there is no $d \in \mathbf{Z}$ which divides all but $c|W_2|$ elements of W_2.

Set $W_1 := k^{-1}\cdot V_1$ (in $\mathbf{F}_p$). Then $V = k\cdot(W_1 \cup W_2) := k \cdot W$, we have

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

We split the sum into two parts,

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

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

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

Lemma 3.9.1.

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

Proof. Making use of (3.13), we split the sum into three subsums, according to the magnitude of $\|\xi\|$.

Small ξ: $\frac{\log^2 n}{n^{3/2}} \le \|\frac{\xi}{p}\| \le \frac{1}{2Cn}$. Similar to Section 3.6, we easily obtain that

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

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

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

Lemma 3.9.2. *Let $a \in \mathbf{F}_p$ be arbitrary. Let $\xi \in \mathbf{F}_p$ and $l > 0$ such that $l\|\xi\| \le p/2$. Then the following holds for any sequence $0 \le i_1 < \cdots < i_m \le l$ with $m \ge 4$*

$$\sum_{j=1}^{m} \|a + i_j \xi\|^2 \ge \frac{m^3}{48} \|\xi\|^2. \tag{3.28}$$

Proof. (of Lemma 3.9.2). Without loss of generality we assume that m is even. For each $j \le m/2$, by the Cauchy-Schwarz inequality and the triangle inequality,

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

Sum over $0 \le j \le m/2$ and notice that $i_{m-j} - i_j$ decreases in j

$$\sum_{j=1}^{m}\|a+i_j\xi\|^2 \geq \frac{1}{2}\sum_{1\leq j\leq m/2}(i_{m-j}-i_j)^2\|\xi\|^2 \geq \frac{1}{2}\sum_{1\leq j\leq m/2} j^2\|\xi\|^2$$
$$\geq \frac{m^3}{48}\|\xi\|^2.$$

□

Now we return to our main goal. 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 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

$$\frac{1}{2}\frac{|W_2|/2}{l/8C} = \frac{2C|W_2|}{l} \geq \frac{Cn}{l}.$$

Now we apply (3.28) to each block (with $m = \lfloor l/8C\rfloor$), and then sum over the collection of all blocks,

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

where in the last bound we used $l = \frac{1}{2\|\xi/p\|}$.

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

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

Assume, for a moment, that $|W'| \leq (1-c)|W_2|$ 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 we are done.

So it suffices to show that $|W'| \leq (1-c)|W_2|$ 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', which contradicts the c-irreducibility assumption of W_2.

To obtain the above contradiction we use the following lemma, which is a consequence of Lemma 1.1.2.

Lemma 3.9.3. *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 $2\gamma n + 1$.*

Applying Lemma 3.9.3 to W_2, we infer that that $kW_2 - kW_2$ contains an arithmetic progression $Q = \{id : |i| \leq \gamma n\}$. Since $Q \subset [-2kCn, 2kCn]$, the step d must be bounded,

$$0 < d \leq \frac{2kC}{\gamma}. \tag{3.29}$$

Let q be an element of Q. By definition, $q = w_1 + \cdots + w_k - w_{k+1} - \cdots - w_k$ for some $w_i \in W_2$. Since $\|\frac{w\xi}{p}\| \leq \delta$ for all i, by the triangle inequality we have $\|\frac{q\xi}{p}\| \leq 2k\delta$.

We apply the above estimate for all elements of Q, obtaining $\|\frac{id\xi}{p}\| \leq 2k\delta$ for all $|i| \leq \gamma n$. But since $2k\delta < \frac{1}{4}$, it follows that $\|\frac{d\xi}{p}\| \leq \frac{2k\delta}{\gamma n}$.

Next, we view $\mathbf{F}_p$ as the interval $[-(p-1)/2, (p-1)/2]$ of $\mathbf{Z}$, and consider ξ as an integer satisfying $\frac{p}{64C} \leq |\xi| \leq \frac{p}{2}$. By the above bound of $\|\frac{d\xi}{p}\|$, we have $d\xi = sp + t$, where $|t| \leq \frac{2k\delta p}{\gamma n}$.

We write $\xi = \frac{sp+t}{d}$. As ξ has large absolute value, s cannot be zero. Another crucial observation is that as $|\xi| \leq p/2$ and t is small, d does not divide s.

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

$$w\xi = \frac{wsp}{d} + \frac{wt}{d} = s'p + \frac{t'p}{d} + \frac{wt}{d}, \tag{3.30}$$

where $ws = s'd + t'$; $s', t' \in \mathbf{Z}$, and $-d/2 \leq t' \leq d/2$.

Now, since $|w| \leq Cn$, we have $|\frac{wt}{d}| \leq Cn\frac{2k\delta p/\gamma n}{d} = \frac{2Ck\delta}{\gamma}\frac{p}{d} \leq \frac{p}{2d}$, where in the last inequality we used the fact that δ is small compared to all other quantities.

Next we consider two cases, according to the value of t'.

Case 1 : $t' \neq 0$. We have $\frac{p}{d} \leq |\frac{t'p}{d}| \leq \frac{p}{2}$, and so $\frac{p}{2d} \leq |\frac{t'p}{d} + \frac{wt}{d}| \leq \frac{p}{2} + \frac{p}{2d}$.

It is implied that $\|\frac{w\xi}{p}\| \geq \frac{1}{2d}$, and hence, by the bound of d from (3.29), $\|\frac{w\xi}{p}\| \geq \frac{\gamma}{4kC} > \delta$. But this inequality violates the definition of W'.

Case 2: $t' = 0$. It follows that d divides sw for all $w \in W'$. Recall that d does not divide s, we conclude that all the element of W' is divisible by a nontrivial divisor of d, which contradicts the c-irreducibility assumption of W_2. This concludes the proof of Lemma 3.9.1.

□

We have shown that the contribution of Σ_2 is negligible. Thus, it suffices to justify Theorem 3.2.5 from the following assumption

$$\Sigma_1 \geq (1-\epsilon)\sqrt{\frac{24}{\pi}}n^{-3/2} = (1-\epsilon)\sqrt{\frac{12}{\pi^2 n^3}}\int_{-\infty}^{\infty}\exp(-\frac{y^2}{2})dy. \tag{3.31}$$

We have

$$\begin{aligned}\Sigma_1 &= \frac{1}{p}\sum_{\|\frac{\xi}{p}\| \leq \frac{\log^2 n}{n^{3/2}}} e_p(-\frac{v\xi}{2})\prod_{w\in W}\cos\frac{2\pi w\xi}{p} \\ &\leq \frac{1}{p}\sum_{\|\frac{\xi}{p}\| \leq \frac{\log^2 n}{n^{3/2}}}\prod_{w\in W_1}|\cos\frac{\pi w\xi}{p}|\prod_{w\in W_2}\cos\frac{\pi w\xi}{p} \\ &\leq \frac{1}{p}\sum_{\|\frac{\xi}{p}\| \leq \frac{\log^2 n}{n^{3/2}}}\exp\left(-\sum_{w\in W_1}\|\frac{\xi w}{p}\|^2\right)f_{W_2}(\xi),\end{aligned}$$

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

$$\sum_{w \in W_2} w^2 \leq (1+\epsilon)\frac{n^3}{12}. \tag{3.32}$$

To obtain information about W_1, we need to restrict the range of ξ furthermore. Let $C = C(\epsilon)$ be a number so that

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

Notice that if ϵ is sufficiently small, then by the property of the Gaussian distribution

$$\int_{C/2 \leq |y| \leq C} \exp(-\frac{y^2}{2})dy \geq \epsilon^{2/3}. \tag{3.33}$$

Next, as p and n are large

$$\begin{aligned}\frac{1}{p} \sum_{\frac{C}{n^{3/2}} \leq \|\frac{\xi}{p}\| \leq \frac{\log^2 n}{n^{3/2}}} f_{W_2}(\xi) &\leq (\sum_{w \in W_2} \pi^2 w^2)^{-1/2} \int_{|y| \geq C} \exp(-\frac{y^2}{2})dy \\ &< (1+4\epsilon)^{-1}\sqrt{\frac{12}{\pi^2 n^3}}\epsilon \leq \epsilon\sqrt{\frac{12}{\pi^2 n^3}} \int_{-\infty}^{\infty} \exp(-\frac{y^2}{2})dy,\end{aligned}$$

where we used the estimate $\sum_{w \in W_2} w^2 \geq (1-3\epsilon)\frac{n^3}{12}$ as $|W_2| \geq (1-\epsilon)n$.

It follows from (3.31) that

$$\frac{1}{p} \sum_{\|\frac{\xi}{p}\| \leq \frac{C}{n^{3/2}}} \exp\left(-\sum_{w \in W_1} \|\frac{w\xi}{p}\|^2\right) f_{W_2}(\xi) \geq (1-2\epsilon)\sqrt{\frac{12}{\pi^2 n^3}} \int_{-\infty}^{\infty} \exp(-\frac{y^2}{2})dy. \tag{3.34}$$

Viewing $\mathbf{F}_p$ as the interval $[-(p-1)/2, (p-1)/2]$ of $\mathbf{Z}$, we now show that the elements of W_1 do not have large absolute values.

Lemma 3.9.4. *Let w_0 be an arbitrary element of W_1. Then*

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

Proof. (of Lemma 3.9.4) Assume, for a contradiction, that $|w_0| > \frac{n^{3/2}}{2C}$. 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, we then have $|I| \geq Cp/n^{3/2}$. By definition, if $\xi \in I$ then

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

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

$$\begin{aligned}\Sigma_1 &\leq \frac{1}{p}\sum_{\xi \in I} \exp(-\|\frac{w\xi}{p}\|^2) f_{W_2}(\xi) + \frac{1}{p} \sum_{\|\frac{\xi}{p}\| \leq \frac{C}{n^{3/2}}, \xi \notin I} \exp(-\|\frac{w\xi}{p}\|^2) f_{W_2}(\xi) \\ &\leq \frac{1}{p}\sum_{\xi \in I} \exp(-\frac{1}{64}) f_{W_2}(\xi) + \frac{1}{p} \sum_{\|\frac{\xi}{p}\| \leq \frac{C}{n^{3/2}}, \xi \notin I} f_{W_2}(\xi) \\ &\leq \frac{1}{p} \sum_{\|\frac{\xi}{p}\| \leq \frac{C}{n^{3/2}}} f_{W_2}(\xi) - (1 - \exp(-\frac{1}{64}))\frac{1}{p} \sum_{\frac{C}{2n^{3/2}} \leq \|\frac{\xi}{p}\| \leq \frac{C}{n^{3/2}}} f_{W_2}(\xi).\end{aligned}$$

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

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

where in the last estimate we used (3.33).

By choosing ϵ sufficiently small, this upper bound is smaller than the lower bound provided in (3.31), a contradiction. This concludes the proof of Lemma 3.9.4. □

To continue, set

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

As $|w| \leq \frac{n^{3/2}}{2C}$ for all $w \in W_1$ and $\|\frac{\xi}{p}\| \leq \frac{C}{n^{3/2}}$, we have

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

Applying the same argument as in Lemma 3.9.4, we obtain

$$\sigma_1 \leq \frac{n^3}{4C^2}.$$

Recall that $W = W_1 \cup W_2$ and $\sum_{w \in W_2} w^2 \leq (1+\epsilon)n^3/12$ from (3.32), we conclude that

$$\begin{aligned} \sum_{w \in W} w^2 = \sum_{w \in W_1} w^2 + \sum_{w \in W_2} w^2 &\leq (1 + \epsilon + \frac{3}{C^2})\frac{n^3}{12} \\ &\leq (1+\epsilon')\frac{n^3}{12}, \end{aligned}$$

where $\epsilon' \to 0$ as $\epsilon \to 0$, finishing the proof.

References

[1] N. Alon, *Subset sums*, Journal of Number Theory, 27 (1987), 196-205.

[2] N. Alon and G. Freiman, *On sums of subsets of a set of integers*, Combinatorica, 8 (1988), 297-306.

[3] Y. Bilu, *Structure of sets with small sumset*, Structure theory of set addition. Asterisque 258 (1999), xi, 77-108.

[4] M. C. Chang, *Generalized arithmetical progressions and sumsets*, Acta Math. Hungar. 65 (1994), no. 4, 379-388.

[5] P. Erdős, *Some problems and results on combinatorial number theory*, Proc. 1st. China Conference in Combinatorics (1986).

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

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

[8] P. Frankl and Z. Füredi, *Solution of the Littlewood-Offord problem in high dimensions*, Ann. of Math. (2) 128 (1988), no. 2, 259-270.

[9] G. Freiman, *Foundations of a structural theory of set addition*, translated from the Russian, Translations of Mathematical Monographs, Vol 37. American Mathematical Society, Providence, R. I., 1973.

[10] B. Green, *An exposition on triples in Arithmetic progression*,

http://www.dpmms.cam.ac.uk/˜ bjg23/papers/bourgain-roth.pdf.

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

[12] B. Green and T. Tao, *Compressions, convex geometry and the Freiman-Bilu theorem*, Q. J. Math. 57 (2006), no. 4, 495–504

[13] J. Griggs, *Database Security and the Distribution of Subset Sums in* $\mathbf{R}^m$, Graph Theory and Combinatorial Biology, Balatonlelle 1996 , Bolyai Math. Studs. 7 (1999), 223–252.

[14] J. Griggs, J. Lagarias, A. Odlyzko and J. Shearer, *On the tightest packing of sums of vectors*, European J. Combin. 4 (1983), no. 3, 231–236.

[15] G. Halász, *Estimates for the concentration function of combinatorial number theory and probability*, Period. Math. Hungar. 8 (1977), no. 3-4, 197-211.

[16] J. Kahn, J. Komlós and E. Szemerédi, *On the probability that a random* ±1 *matrix is singular*, J. Amer. Math. Soc. 8 (1995), 223-240.

[17] G. Katona, *On a conjecture of Erdős and a stronger form of Sperner's theorem.* Studia Sci. Math. Hungar 1 (1966), 59–63.

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

[19] 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.

[20] H. Iwaniec and E. Kowalski, *Analytic number theory*, American Mathematical Society, Colloquium publications, Volume 53.

[21] E. Lipkin *On representation of* $r-$*powers by subset sums*, Acta Arithmetica 52 (1989), 114-130.

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

[23] I. Ruzsa, *An analogue of Freiman's theorem in group*, Structure theory of set addition. Asterisque 258 (1999), 323-326.

[24] M. Rudelson and R. Vershynin, *The Littlewood-Offord problem and the condition number of random matrices*, Advances in Mathematics 218 (2008), no 2, 600-633.

[25] A. Sárközy, *Finite addition theorems I*, J. Num. Thy. 32 (1989), 114–130.

[26] A. Sárközy, *Finite addition theorems, II*, Journal of Number Theory, 48 (1994), 197-218.

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

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

[29] E. Szemerédi and V. H. Vu , *Long arithmetic progression in sumsets and the number of x-free sets.* Proceeding of London Math Society, 90(2005) 273-296.

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

[31] E. Szemerédi and V. Vu, *Finite and Infinite Arithmetic Progressions in Sumsets*, Annals of Mathematics, 163 (2006), no 1, 1-35.

[32] T. Tao, *Freiman's theorem in solvable groups*, `http://arxiv.org/abs/0906.3535`

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

[34] T. Tao and V. Vu, *John-type theorems for generalized arithmetic progressions and iterated sumsets,* Adv. Math. 219 (2008), no. 2, 428–449.

[35] T. Tao and V. Vu, *Random matrices: The Circular Law*, Communication in Contemporary Mathematics 10 (2008), 261-307.

[36] T. Tao and V. Vu, *From the Littlewood-Offord problem to the circular law: universality of the spectral distribution of random matrices,* Bull. Amer. Math. Soc. (N.S.) 46 (2009), no. 3, 377–396.

[37] 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.

[38] T. Tao and V. Vu, *A sharp inverse Littlewood-Offord theorem*, to appear in Random Structures and Algorithms.

[39] T. Tao and V. Vu, *Smooth analysis of the condition number and the least singular value*, (to appear in Mathematics of Computation).

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

Vita

Hoi H. Nguyen

2005-2010	Ph. D. in Mathematics, Rutgers University
2000-2005	M. S. in Mathematics, Eötvös Loránd University, Hungary
1998	Graduated from Le Khiet high school, Vietnam

2005-2010	Teaching assistant, Department of Mathematics, Rutgers University