Approximate Groups in Additive Combinatorics: A Review of Methods and Literature

Lloyd Husbands

A dissertation submitted to The University of Bristol in accordance with the requirements for award of degree of MSc. in the Faculty of Science, Department of Mathematics, September, 2009.

Thirty-one thousand, five-hundred and twenty words.

Abstract

Herein lies a discussion of a relatively new concept of 'approximate grouplikeness' of subsets of additive combinatorics. There is a thorough discussion of how the now classical problem of small doubling in the integers fits into this framework, and a detailed analysis of known results in this case and in sets of residues modulo a prime. During this discussion, an improvement over Rodseth's version of Freiman's 2.4-Theorem is presented. Later, there is a translation and modernisation of a Russian paper of Freiman which amounts, as far as the author is aware, to pretty much all that is known in the general setting about finite sets with small doubling in arbitrary groups.

The last section describes the behaviour of the algebra norm and its worthiness as a measure of grouplikeness, and an improvement to a four-decade old paper of Saeki is presented.

Contents

1. Introduction

Broadly speaking, additive combinatorics is the study of additive behaviour in abelian groups. The area is home to many elegant, famous theorems and fascinating qualitative results concerning many different phenomena. Intuitively, making strong statements about additive structure of objects necessitates strong structural consequences, and it is from this intuition that the area draws motivation for many of its most famous results. The area incorporates elements of many other areas of maths in its proofs, most notably combinatorics, classical number theory, and Fourier analysis.

To illustrate some of the themes common to much of additive combinatorics, it is best to start off with the example of Freiman's theorem. To allow a precise statement of the theorem, two notions need be introduced. The first is that of set addition: for two subsets A, B of an abelian group G, $A + B$ is defined to be the set of all sums $a + b$, where $a \in A, b \in B$, and we refer to $A + A$ as the sumset of A. The second is that of a generalised arithmetic progression: a generalised arithmetic progression P is a set $P = A_1 + A_2 + \cdots + A_n$, where each A_i $(1 \leqslant i \leqslant n)$ is itself an arithmetic progression, and the rank of P is the least n needed to write P in this way.

Also, we let $|A|$ denote the size of a finite set A.

Theorem 1.0.1 (Freiman's Theorem)**.** *Let $K > 1$ be a real number. If A is a set of integers such that $|A + A| < K|A|$, then there is an a generalised arithmetic progression P of size n and rank r containing A, where n and r depend only on K.*

Freiman's theorem is the prototypical example of additive combinatorics. It is precise, yet it simultaneously masks the burden of imprecision, a problem inherent throughout the entire subject. In vague terms, Freiman's theorem asserts that a set of integers that grows in size linearly under auto-addition 'looks like' an arithmetic progression, but what 'looks like' has insofar been proven to mean is worryingly meagre.

The theorem infers the existence of functions $n(K)$ and $r(K)$, namely the least such n and r, respectively, for which the theorem holds. Indeed, the smaller $n(K)$ and $r(K)$ are for any given K, then the more A looks like a generalised arithmetic progression. However, the known bounds for these functions are quite large. There have been several quantitative proofs of Freiman's theorem, but as far as the author is aware the best known bounds are due to Bilu, and has $r(K)$ linear in K, while $n(K)$ is exponential in K.

So, as is typical throughout additive combinatorics, Freiman's theorem appealingly infers a lot of structure from very little additive information, yet the achievable bounds for the theorem are disappointly impracticable. The appeal of the subject lies in the fact that such a powerful statement is actually proven true, and now it a widely-regarded common goal to prove that one may bound $n(K)$ and $r(K)$ by polynomials in K. This statement is itself known as the Polynomial Freiman-Ruzsa Conjecture and, if proven, would represent a milestone within the area, though one should note it may be the case that the statement of Freiman's theorem need be altered slightly for this to be true.

One may view Freiman's theorem in the more general framework that will be here addressed. Firstly, one need not be restricted solely to the integers, as the statement extends in a very natural way into any abelian group. Indeed, sensible statements may even be formulated in non-abelian groups, though the statements there can appear rather more intimidating, as we shall later see.

Thus our first reconsideration of Freiman's theorem shall be to rephrase the statement in other groups, primarily in residues modulo a prime. In this instance, the integers provide a useful sandpit for considering results because residues modulo a prime resemble the integers, in a strong additive sense that we shall later make precise. In fact, much of the proofs we will present will lend much to the idea of replacing residues entirely by a system of integers that behave identically to sets of residues additively, a method known as rectification.

Replacing a statement in the residues by a statement in the integers is useful to us because we can then use Freiman's theorem to infer a structural theorem for sets of residues with small sumset. Interestingly, the more accessible proofs of Freiman's theorem work in rather the opposite direction, but the mechanic itself works either way so long as a structural statement is known at one end of the rectification process.

Rectification is a method that enables us to reconsider addition in one group as addition in another group where we know more about how addition works, and it works by finding a Freiman isomorphismic image between two sets. Suppose we have two groups G, G', a subset $A \subseteq G$, and a function $\varphi : G \to G'$. If it's the case that whenever $a + b = c + d$ $(a, b, c, d \in A)$ we also have $\varphi(a) + \varphi(b) = \varphi(c) + \varphi(d)$, then we call φ a Freiman homomorphism. If there exists a Freiman

homomorphism from a set $A \subseteq G$ to a set $A' \subseteq G'$, we say that A is Freiman homomorphic to A'. Further, if A' is Freiman homomorphic to A, we say that A and A' are Freiman isomorphic.

Freiman isomorphisms are useful because, as far as sumsets are considered, we have identified the sets A and $2A$ in G with sets A' and $2A'$ in G', and any structural statement we prove for the set A' tells us about the structure of A via the Freiman isomorphism.

We shall also consider other groups where we will not be able to refer to our result in the integers, so other combinatorial methods are required to resolve statements resolving Freiman's theorem. The statement of Freiman's theorem above is a formulation exclusive to the integers. Other groups require slightly different notions of structure than just a generalised arithmetic progression, and a very general formulation of Freiman's theorem is known in any finite abelian group, a result due to Green and Ruzsa. Thus, in some sense, Freiman's theorem is resolved, and we consider putting the scheme of consideration of Freiman's theorem into a more general framework.

So we now consider A as a finite subset of an abelian group G. In this situation, there is a more general version of Freiman's theorem to replace Theorem 1.0.1. In any case, it is easy to show that $|A+A| = |A|$ is attainable if and only if A is a coset of a subgroup of G. Thus one might consider the requirement $|A + A| < K|A|$ to be demanding that A behave approximately like a group. One can make this notion precise by defining the doubling constant $\sigma(A)$ of a set A to be

$$\sigma(A) = \frac{|2A|}{|A|}.$$

In this manner, $\sigma(A)$ is a statistic that measures how close A is to behaving like a group and, properly modified, Freiman's theorem is confirmation that the smaller this statistic is, the more A looks like a subgroup of G.

Of course, one can imagine other statistics that measure, in some capacity, how like a group a given set A is. One that we shall discuss, related to the Littlewood problem, is the algebra norm $\|A\|$, defined by

$$\|A\| := \sum_{\gamma \in \widehat{G}} |\widehat{1_A}(\gamma)|.$$

Here, $\widehat{f}$ indicates the Fourier transform of a function f. There is a straightforward proof that $\|A\| \geqslant 1$ for all sets A, with equality if and only if A is a coset of a subgroup.

In spirit, Cohen's theorem is to the algebra norm as Freiman's theorem is to the doubling constant; it tells us that the closer the algebra norm of A is to this minimal value, the more closely A mimics the behaviour of a subgroup of G, though it should be noted that for our consideration it shall be more illustrative to consider a recent theorem of Green and Sanders rather than that of Cohen, as we shall only be considering finite groups G for this question, where Cohen's theorem is vacuous.

Furthermore, we shall only be interested in cases where these measures of being 'approximately grouplike' parameters are tiny. As was mentioned earlier, Freiman's theorem is impractical in the sense that the information it offers is, in an obvious sense, quite weak. There are cases, however, in which strong assertions may be made. If one

considers our Freiman theorem earlier in the case where $K = 2$, for example, then it is an elementary fact that A itself must be an arithmetic progression. Similarly, there is the following result, due to Freiman.

Theorem 1.0.2 (Freiman's $3k-3$ Theorem)**.** *If A is a set of integers such that $|A+A| < 3|A| - 3$, then A is a subset of an arithmetic progression P of size $|P| \leqslant |2A| - |A| + 1$.*

So, if one considers cases where $\sigma(A)$ is particularly small, very precise results can be attained. Similarly, when considering the algebra norm, the bounds given by the Green-Sanders theorem are at least doubly-exponential in $\|A\|$, whereas if one considers the case where $\|A\| < 9/7$ then one can get a surprisingly precise answer.

With this in mind, we shall discuss these two different measures of how subsets of abelian groups may behave approximately like subgroups, and in particular how stable these measures are when varying slightly from the minimal value. The results we shall see are satisfyingly succinct.

The first section shall introduce notation that we shall use throughout, as well as some basic results that are either essential or motivate later discussion. The subsequent chapters shall discuss efforts to extend results in the integers to other groups, as we described above. We include a translation of a paper of Freiman which considers a formulation of Freiman's theorem in non-abelian groups, for the small-parameter case. This paper could amount to pretty much all that is known about Freiman's theorem in the non-abelian case: it is not even clear what the correct version of Freiman's theorem would look like in a general group, so that there is an answer for this particular case provides justification for the study of small-parameter versions in general, and also

provides a vague hope that a full statement of Freiman's theorem may be attainable in some form in non-abelian groups.

We shall present many ideas and arguments that fit into this framework and mention where proofs come from. In most cases, we present simplified versions of proofs found in the literature. In doing so, we discuss briefly the papers from which the proof comes and how our proof is simpler, or not — in a couple of cases, the proofs are no simpler but are presented because they are inkeeping with the theme of our discussion or are, to put it simply, elegant. Sometimes, the proofs we present are much simpler because they are special cases of more general theorems, and we hope that our simplified versions serve to elucidate and explain the original proof and put them into the context of our discussion. Thus our discussion can be considered a review of the literature in the area of our study. When there are papers related to our discussion but we do not go into them, we make note of them for the interested reader.

In some instances the proofs will be new creations of the author and, though these instances are rare, we shall try to make note of them.

2. Notation and Basic Results

This section introduces the notation that we shall be using throughout our discussions. Our main tools will be combinatorics and Fourier analysis, so this chapter will more clearly state the definitions of set addition and Fourier transform that were alluded to in the introduction. Afterwards, although it may merely be revision for the initiated reader, we go over a few results of Fourier analysis for finite groups to make our later discussion more straightforward.

For now we shall only consider abelian groups, though it should be noted that some of the later sections will cover some non-abelian cases. The reader should notice how the definitions presented here are inadequate for that purpose.

So, for the benefit of the following few sections, we shall consider G to be an abelian group.

2.1. **Set Addition.** Our first task is to extend the definition of addition from mere elements of the group to subsets of the group. There is an obvious and straightforward way to do this. Firstly, we define $-A = \{-a : a \in A\}$ for all sets $A \subseteq G$. Then we define $A + B$ as

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

that is, $A + B$ consists of all elements of G that can be written as a sum of an element of A and an element of B. We similarly define $A - B := A + (-B)$ straightforwardly.

The associativity of addition immediately infers associativity in set addition, so there is no ambiguity, for example, if we write $A + B + C$ to represent the set $(A + B) + C$. Thus we can now add and subtract sets from one another and the order in which we write the summands

makes no difference: just what we want from an abelian system of set addition.

Note, however, that this does not turn the subsets of G into a group under set addition. Indeed, were it a group, the zero element would have to be $\{0_G\}$, where 0_G is the neutral element of G. However, it is rarely true, for instance, that $A - A = \{0_G\}$ – this holds if and only if A is a singleton – so do not think of set addition as being a group operation.

We shall often talk of *translates* of a set $A \subseteq G$, by which we mean a set $\{g\} + A$ for some $g \in G$. It shall often be convenient to omit the curly-brackets, so we let $g + A$ denote the set $\{g\} + A$ as a matter of definition.

Sometimes, we shall want to talk of *dilates* of a set $A \subseteq G$, by which we mean a set $\{xa : a \in A\}$ for some $x \in G$; we shall let $x \times A$ denote this set. Note that, in particular, if $x \in \mathbb{Z}$, then the two sets $x \times A$ and xA are almost always different sets, and should not be confused with one another.

2.2. **Set Addition and Convolution.** To discuss sizes of sets in a more systematic way, we introduce indicator functions to count the sizes of sets. For a set $A \subseteq G$, the function $1_A : G \to \{0, 1\}$ defined by

$$1_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{otherwise,} \end{cases}$$

for all $x \in G$, is called the indicator function of A, and is defined so that $\sum_{x \in G} 1_A(x) = |A|$.

It shall be convenient to discuss the indicator function of $A + A$ in terms of the indicator function of A, and this in done by convolving the

indicator function of A with itself. The convolution of two functions $f, g : G \to \mathbb{C}$ is the function $f * g : G \to \mathbb{C}$ given by

$$f * g(x) := \mathop{\mathbb{E}}_{y \in G} f(y)g(x-y).$$

A simple substitution of summands in the definition above shows that $f * g = g * f$ for all functions f, g. If we define the support $\mathrm{supp}(f)$ of a function $f : G \to \mathbb{C}$ to be

$$\mathrm{supp}(f) = \{x \in G : \ f(x) \neq 0\}$$

then it is clear from the definition of convolution that $\mathrm{supp}(f * g) \subseteq \mathrm{supp}(f) + \mathrm{supp}(g)$. In particular, if we consider the convolution of two indicator functions $1_A, 1_B$ of two sets A, B respectively, then more is true: firstly, notice that for any $x \in G$, $p \times 1_A * 1_B(x)$ is the number of ways of writing x is a sum of elements $a + b$, where $a \in A$ and $b \in B$. What is more, this means, we in fact have

$$\mathrm{supp}(1_A * 1_B) = \mathrm{supp}(1_A + 1_B) = A + B.$$

As we shall explain later, it is this relation that makes Fourier analysis so useful in discussing set addition.

2.3. **Characters on Finite Abelian Groups.** G being abelian allows us to define a *character* on G to be precisely a homomorphism $\gamma : G \to \mathbb{C}^\times$, where $\mathbb{C}^\times$ denotes the group of non-zero complex numbers under multiplication.

The set of all charaters on G is called the dual group of G, and denoted $\widehat{G}$. We turn $\widehat{G}$ into a multiplicative abelian group by defining $\gamma_1\gamma_2(g) = \gamma_1(g)\gamma_2(g)$ for all $g \in G, \gamma_1, \gamma_2 \in \widehat{G}$.

Thus we now have reason to view abelian groups from both an additive and multiplicative perspective, thus it shall prove useful to make two further definitions. Let N be a positive integer. $\mathbb{Z}/N\mathbb{Z}$ shall henceforth denote the additive abelian group of integers modulo N, and U_N shall denote the multiplicative abelian group of N-th roots of unity.

Be it from an additive or multiplicative viewpoint, recall the following *Fundamental Theorem of Finite Abelian Groups:* Every finite abelian group is isomorphic to a direct sum of cyclic groups.

We shall now prove that $G \cong \widehat{G}$. By the Fundamental Theorem of Finite Abelian Groups, $G \cong \oplus_{j=1}^{k}(\mathbb{Z}/n_j\mathbb{Z})$ for some positive integers $n_1, \ldots, n_k$. So if we let $e_j = (0, \ldots, 0, 1, 0, \ldots, 0)$ be the element that is non-zero in the jth position, then $\{e_1, \ldots, e_k\}$ is a basis for G.

Now consider an arbitrary character γ on G. As the generator e_j has order n_j, we must have $\gamma(e_j)^{n_j} = 1$; that is $\gamma(e_j)$ is a n_jth root of unity for each j. Furthermore, the values of γ are determined by the relation

$$\gamma(i_1e_1 + \cdots + i_ke_k) = \gamma(e_1)^{i_1} \cdots \gamma(e_k)^{i_k}$$

so we can identify γ with the element $(\gamma(e_1), \ldots, \gamma(e_k)) \in \oplus_{j=1}^{k} U_{n_j}$ and this identification is a homomorphism. Conversely, given any element $(\lambda_1, \ldots, \lambda_k) \in \oplus_{j=1}^{k} U_{n_j}$, we can define a corresponding character γ given by

$$\gamma(i_1e_1 + \cdots + i_ke_k) = \lambda_1^{i_1} \cdots \lambda_k^{i_k}.$$

Thus we have the isomorphisms

$$\widehat{G} \cong \oplus_{j=1}^{k} U_{n_j} \cong \oplus_{j=1}^{k}(\mathbb{Z}/n_j\mathbb{Z}) \cong G,$$

or, more concisely, $G \cong \widehat{G}$.

Given this duality, it might seem intuitive to consider G and $\widehat{G}$ to be the same object. From the point of view of Fourier analysis, this actually turns out to be unhelpful. We shall want G and $\widehat{G}$ to satisfy the so-called *Fourier inversion formula*, and this requires that G and $\widehat{G}$ be given different measures.

For a finite group G, the function $\mu_G : G \to [0,1]$ defined by $\mu_G(A) = |A|/|G|$ is called the *Haar measure* of G. On the other hand, the function $m_G : G \to \mathbb{N}$ defined by $m_G(A) = |A|$ is called the *counting measure* of G. Integration with respect to these measures has $d\mu_G(x) = 1/|G|$ and $dm_G(x) = 1$ respectively.

This allows us to discuss the orthogonality properties of characters. Precisely, these are the statements

$$\int_G \gamma(x)\, d\mu_G(x) = \begin{cases} 1 & \text{if } \gamma \text{ is constant} \\ 0 & \text{otherwise;} \end{cases}$$

and

$$\int_{\widehat{G}} \gamma(x)\, dm_{\widehat{G}}(\gamma) = \begin{cases} |G| & \text{if } x = 0_G \\ 0 & \text{otherwise.} \end{cases}$$

They are both proved trivially. For instance, in proving the first, fix an arbitrary $g \in G$. Then we have

$$\int_{x\in G} \gamma(x)\, dm_G(x) = \int_{x\in G} \gamma(x+g)\, dm_G(x) = \gamma(g) \int_{x\in G} \gamma(x)\, dm_G(x),$$

so either $\gamma(g) = 1$ for all $g \in G$ or the integral is zero. In proving second, all that is needed is the observation that the elements of G are essentially characters on $\widehat{G}$ by defining $x(\gamma) = \gamma(x)$ for each $x \in G, \gamma \in \widehat{G}$.

2.4. **Expectation Notation.** We shall be make use of these orthogonality relations shortly, but before that we remark on our notation.

An area of thinking that has proved incredibly useful in additive combinatorics is the probabalistic method. Viewing integrals over G with respect to the Haar measure as averages over G makes application of the probablistic method especially transparent. It has therefore proven useful to use expectation notation to present these integrals. So, for a function $f : G \to \mathbb{C}$, we define

$$\mathbb{E}_G f := \int_G f \, d\mu_G = \frac{1}{|G|} \sum_{g \in G} f(g).$$

Similarly, when integrating over $\widehat{G}$, it is practical to consider this as summation because

$$\sum_{\widehat{G}} f = \int_{\widehat{G}} f \, dm_{\widehat{G}}$$

so we shall we use summation notation when doing so. We shall not see many probabilistic arguments, but it is good practice to view our integrals in this way if one plans further study in additive combinatorics.

This expectation notation has the further practical use of distinguishing more clearly where integrals are taking place. Throughout our applications, there will be integrals taking place in G alongside integrals taking place in $\widehat{G}$. That the respective domains of these integrals are separate is clearly shown by the distinction between expectation notation and summation, and proves very useful in understanding, motivating and generalising our arguments.

2.5. **The Fourier Transform on Finite Abelian Groups.** Now, we move onto introducing the Fourier transform for finite abelian groups.

The reader may be familiar with the Fourier transform on $\mathbb{Z}$ or $\mathbb{R}$, but a similar scheme of Fourier analysis may be enacted upon other groups, although many of the results we discuss are directly analagous to those commonly cited for classical Fourier analysis.

Indeed, Rudin [R1] argues that the domain of Fourier analysis is that of the locally compact abelian groups, and Fourier analysis behaves very consistently there. Many of the technicalities Rudin discusses are straightforwardly resolved in finite abelian groups, making our discussion of Fourier analysis relatively simple.

For each function $f : G \to \mathbb{C}$, we define the Fourier transform of f to be the function $\widehat{f} : \widehat{G} \to \mathbb{C}$ defined by

$$\widehat{f}(\gamma) := \mathop{\mathbb{E}}_{g \in G} f(g)\gamma(g) = \frac{1}{|G|} \sum_{g \in G} f(g)\gamma(g)$$

for all $\gamma \in \widehat{G}$.

The Fourier transform has some a couple of nice properties, each stemming from the orthogonality of characters mentioned earlier. The first is the so-called inversion formula, which we can summarise as

$$f(g) = \sum_{\gamma \in \widehat{G}} \widehat{f}(\gamma)\overline{\gamma(g)}$$

for all $g \in G$. This is a direct consequence of the orthogonality relations. To show this, we need only substitute for the definition of the

Fourier transform, as follows:

$$\begin{aligned}\sum_{\gamma\in\widehat{G}} \widehat{f}(\gamma)\overline{\gamma(g)} &= \sum_{\gamma\in\widehat{G}} \mathop{\mathbb{E}}_{x\in G} f(x)\gamma(x)\overline{\gamma(g)} \\ &= \sum_{\gamma\in\widehat{G}} \mathop{\mathbb{E}}_{x\in G} f(x)\gamma(x-g) \\ &= \mathop{\mathbb{E}}_{x\in G} f(x) \sum_{\gamma\in\widehat{G}} \gamma(x-g)\end{aligned}$$

.

By the orthogonality relations, this rightmost sum is zero, except when $x = g$, in which case it takes the value $|G|$, and thus the inversion formula is proven. The second interesting relation is Parseval's identity, which we state as

$$\mathop{\mathbb{E}}_{x\in G} f(x)\overline{g(x)} = \sum_{\gamma\in\widehat{G}} \widehat{f}(\gamma)\overline{\widehat{g}(\gamma)}.$$

for all functions $f, g : G \to \mathbb{C}$. This time, the proof relies only on the inversion formula just proven, but we go through the details to familiarise the reader with the technique. Indeed, we have

$$\begin{aligned}\sum_{\gamma\in\widehat{G}} \widehat{f}(\gamma)\overline{\widehat{g}(\gamma)} &= \sum_{\gamma\in\widehat{G}} \widehat{f}(\gamma)\overline{\mathop{\mathbb{E}}_{x\in G} g(x)\gamma(x)} \\ &= \sum_{\gamma\in\widehat{G}} \widehat{f}(\gamma) \mathop{\mathbb{E}}_{x\in G} \overline{g(x)\gamma(x)} \\ &= \mathop{\mathbb{E}}_{x\in G} \overline{g(x)} \sum_{\gamma\in\widehat{G}} \widehat{f}(\gamma)\overline{\gamma(x)} \\ &= \mathop{\mathbb{E}}_{x\in G} \overline{g(x)} f(x).\end{aligned}$$

2.6. **Convolution and the Fourier Transform.** The power behind Fourier analysis when considering set addition comes from the convenient behaviour of the Fourier transform of convolutions. More precisely, set addition is closely related to the convolution of functions, and convolutions are easily handled by the Fourier transform.

This allegedly convenient behaviour is often summarised by the statement that the 'Fourier transform turns convolution into multiplication'. What is meant by this is $\widehat{f * g} = \widehat{f} \cdot \widehat{g}$; that is, for all $\gamma \in \widehat{G}$,

$$\widehat{f * g}(\gamma) = \widehat{f}(\gamma)\widehat{g}(\gamma).$$

The proof is almost a matter of definition:

$$\begin{aligned}
\widehat{f}(\gamma)\widehat{g}(\gamma) &= \mathop{\mathbb{E}}_{x\in G}\mathop{\mathbb{E}}_{y\in G} f(x)\gamma(x)g(y)\gamma(y) \\
&= \mathop{\mathbb{E}}_{x\in G}\mathop{\mathbb{E}}_{y\in G} f(x-y)\gamma(x-y)g(y)\gamma(y) \\
&= \mathop{\mathbb{E}}_{x\in G}\gamma(x)\mathop{\mathbb{E}}_{y\in G} f(x-y)g(y) \\
&= \mathop{\mathbb{E}}_{x\in G}\gamma(x)(f*g)(x) = \widehat{f*g}(\gamma).
\end{aligned}$$

Multiplication is a far simpler operation to consider than convolution, so this equality is a blessing. Furthermore, if one recalls our discussion of how set addition and convolution are related, one might already be able to see the potential for utilising the Fourier transform as a tool for discussing set addition. If not, our later applications shall be seen to justify our claim.

3. Auto-Addition and Inverse Theorems in the Integers

When considering set auto-addition, there are a couple of questions that immediately crop up. Informally, if one considers auto-addition to be an action on a set A, then one can ask how small and how large (in terms of $|A|$) a set can become under auto-addition, and elementary arguments will give us a lower and upper bound, respectively, in each case. Also, we may be able to use additional group-specific information, depending upon the group under consideration, to increase or decrease these bounds. For example, if we had a set $A \subseteq \mathbb{F}_2^n$, then the fact that $a + a = 0_{\mathbb{F}_2^n}$ for all $a \in A$ limits somewhat the possible size of any set A under auto-addition.

First, we shall concern ourselves with the lower bound of sumsets in the integers.

Lemma 3.0.1. *Let A be a finite set of integers. Then $|2A| \geqslant 2|A| - 1$. Furthermore, $|2A| = 2|A| - 1$ if and only if A is an arithmetic progression.*

Proof. Label the elements $a_1, \dots a_n \in A$ in increasing order so that $a_i < a_j$ whenever $i < j$. Firstly, If A is a singleton, then the result is obviously true, so assume that $n := |A| \geqslant 2$. Then we can list $2n - 1$ distinct elements of $2A$ as follows:

$$\begin{aligned} a_1 + a_1 <& a_2 + a_1 < \cdots < a_n + a_1 \\ &< a_n + a_2 < a_n + a_3 < \cdots < a_n + a_n. \end{aligned}$$

So it is certainly true that $|2A| \geqslant 2|A| - 1$, thus satisfying our lower bound. We can list $2n - 1$ distinct elements of $2A$ in another way, as

follows:

$$\begin{aligned} a_1 + a_1 &< a_1 + a_2 \\ &< a_2 + a_2 < a_3 + a_2 < \cdots < a_n + a_2 \\ &< a_n + a_2 < a_n + a_3 < \cdots < a_n + a_n. \end{aligned}$$

As $|2A| = 2|A| - 1$, these two lists must coincide, so we have $2n - 1$ equalities by pairing up the elements in the respective lists. In particular, if we consider the 3rd through nth of these equalities, we find:

$$\begin{aligned} a_3 + a_1 = a_2 + a_2 \quad &\Longrightarrow \quad a_3 = a_2 + (a_2 - a_1); \\ a_4 + a_1 = a_3 + a_2 \quad &\Longrightarrow \quad a_4 = a_3 + (a_2 - a_1); \\ \vdots \qquad\qquad &\qquad\quad \vdots \\ a_n + a_1 = a_{n-1} + a_2 \quad &\Longrightarrow \quad a_n = a_{n-1} + (a_2 - a_1). \end{aligned}$$

Thus if we define $d = a_2 - a_1$, then A is an arithmetic progression with starting term a_1 and common difference d. □

This is our first example of an inverse theorem, but we shall talk more about this later. For now, we discuss the corresponding result for upper bounds.

Lemma 3.0.2. *Let A be a finite set of integers. Then $|2A| \leqslant \binom{|A|+1}{2}$. Furthermore, there exist sets such that $|2A| = \binom{|A|+1}{2}$.*

Proof. Let $n := |A|$ and let $A = \{a_1, \ldots, a_n\}$. First the sums $a_i + a_j$, where $i \neq j$, are duplicated in the form $a_i + a_j = a_j + a_i$, so such pairs contribute at most $\binom{n}{2}$ elements to $2A$. The pairs where $i = j$ can then

contribute at most n further elements, giving us at most

$$\binom{n}{2} + n = \binom{n+1}{2}$$

elements in $2A$, as required. For the final claim, note that $\{1, 3, 3^2, \ldots, 3^{n-1}\}$ is such a set. □

Remark 3.0.3. A set of integers A with the property that $2A$ achieves this maximum size are known as Sidon sets, and we won't discuss these very much. On the other hand, we shall be more interested in sets that have rather smaller sumsets.

So, Lemma 3.0.1 tells us that sets that achieve the smallest possible sumset must have a particular structure. That is called an *inverse theorem*. In general, an inverse theorem supposes statistical information about a set and concludes structural information about the same set. Going from structural information to statistical information is usually more straightforward, so inverse theorems are so-called because they work in the rather less intuitive direction. Most of the results we are interested in are inverse theorems.

The above results concern only a single set A, but it shall be useful to list an asymmetric analogue of 3.0.1. While our interest typically lies in symmetric results (that is, concerning only a single set A), in order to prove these we will sometimes need to consider asymmetric versions of the result (that is, concerning two sets A and B). The following result should illustrate what we mean.

Lemma 3.0.4. *Let A and B be finite sets of integers. Then $|A+B| \geqslant |A|+|B|-1$. Furthermore, if $|A+B| = |A|+|B|-1$, then both A and B are arithmetic progressions with the same common difference.*

Proof. The proof would work almost verbatim from the original, although we approach the first part of the proof differently to show how having two sets under consideration allows us an argument that would be somewhat clumsy otherwise.

Firstly, we know that translating either of the sets A or B will not affect the size of the sumset $A + B$, so we may assume without loss of generality that the largest element of A is zero, and also that the smallest element of B is zero. Then $A+B$ contains the two sets $A+0$, made up of negative elements, and $0+B$, made up of positive elements. These two sets overlap solely in the set $\{0\}$, so $|A+B| \geqslant |A|+|B|-1$.

In order to prove the second claim, assume $A = \{a_1, \ldots, a_n\}$ and $B = \{b_1, \ldots, b_m\}$, where the sequences $(a_i)_{i=1}^n$ and $(b_i)_{i=1}^m$ are strictly increasing. If $|A + B| = |A| + |B| - 1$, then just as in the proof of Lemma 3.0.1 one can collate the two lists

$$a_1 + b_1 < a_2 + b_1 < \ldots < a_n + b_1 < a_n + b_2 < \cdots < a_n + b_m$$

and

$$\begin{aligned} a_1 + b_1 &< a_2 + b_1 \\ &< a_2 + b_3 < a_2 + b_4 < \cdots < a_2 + b_{m-1} \\ &< a_3 + b_{m-1} < a_4 + b_{m-1} < \cdots < a_n + b_{m-1} < a_n + b_m \end{aligned}$$

to find that both A and B are arithmetic progressions with common difference $a_2 - a_1$. Indeed, one can come to the same conclusion by using any two such lists of sums in $A + B$ that do not share a sum, though the above lists make short work of it. □

Notice that the asymmetric analogue of Lemma 3.0.2 offers little information, and requires no proof.

Lemma 3.0.5. *Let A and B be finite sets of integers. Then $|A+B| \leqslant |A| \times |B|$. Moreover, there are sets that achieve this upper bound.*

So it's not always the case that asymmetric analogues are useful to consider.

3.1. **Freiman's $3k-3$ Theorem.** Now we move on one step up in the question of small doubling. So far we have considered the special case $|2A| < 2|A|$ and come up with a very precise answer. Now we shall consider $|2A| < 3|A| - 3$, where we can get another reasonably precise answer. Already, though, we shall see the tell-tale signature of the Freiman theorem inaccuracies. Whereas before, our theorem had both necessity and sufficiency in it's statement, our statement in the following case only has necessity.

Theorem 3.1.1 (Freiman's $3k-3$ Theorem)**.** *Let $A \subseteq \mathbb{Z}$ be a finite set of at least three integers. If $|2A| < 3|A| - 3$ then A is contained in an arithmetic progression of length at most $|2A| - |A| + 1$.*

We present a simplified version of the proof found in [F], by introducing concepts of 'diameter' and 'common difference' of a set: for a set of integers A, we let $\mathrm{diam}(A)$ denote the length of the shortest arithmetic progression containing A; by the 'common difference of A', we mean the positive common difference of the shortest arithmetic progression containing A.

In the proof that follows, it is quite standard to assume that $\min(A) = 0$. We will not do this, however, because it subtly masks the process of the proof, and our introduction of the definition of diameter allows us to do without it.

The proof of Theorem 3.1.1 will be done in two parts.

Lemma 3.1.2. *Let A be a finite set of at least three integers. If* $\operatorname{diam}(A) \leqslant 2|A| - 2$ *then* $|2A| \geqslant |A| + \operatorname{diam}(A) - 1$.

Proof. Firstly, we let P be the shortest arithmetic progression containing A. Secondly, let $S = (\min(A) + A) \cup (\max(A) + A) \subseteq 2A$. That is, S is the set of $2|A| - 1$ 'obvious sums' that we know to be in $A + A$. We need to find a further $\operatorname{diam}(A) - |A|$ elements in $A + A$ to complete the proof. Coincidentally, this is as many elements as there are $P \setminus A$, so that turns out to be a good place to start looking.

So take some $x \in P \setminus A$, and let $I = x + P$, an arithmetic progression of length $\operatorname{diam}(A)$. Then the set I contains $|A| - 1$ elements of S, the obvious sums from A — starting with the smallest element $a' \in A$ that larger than x, and ending with $\max(A) + a''$, where $a'' \in A$ is the largest element of A smaller than x. However, I also contains the set $x + \min(A) + \max(A) - A$, and as P has length $\operatorname{diam}(A) \leqslant 2|A| - 2$, these two sets S and $x + \min(A) + \max(A) - A$ coincide.

S is made up of two simple sets, so either $(\min(A)+A) \cap (x+\min(A)+\max(A) - A) \neq \emptyset$ or $(\max(A) + A) \cap (x + \min(A) + \max(A) - A) \neq \emptyset$. Consequently, at least one of either $x + \min(A)$ or $x + \max(A)$ is in $2A$, neither of which is an element of S. Thus for every $x \in P \setminus A$, we find a new element of $2A$, and hence there are $\operatorname{diam}(A) - |A|$ further elements in $2A$, as we required. □

Lemma 3.1.3. *Let A be a finite set of at least three integers. If* $\operatorname{diam}(A) \geqslant 2|A| - 2$ *then* $|2A| \geqslant 3|A| - 3$.

Proof. When $\operatorname{diam}(A) = 2|A| - 2$, the result follows by the previous theorem, so we may assume that $\operatorname{diam}(A) \geqslant 2|A| - 1$. The rest of our

proof will be an induction on $|A|$ – we skip the straightforward check for $|A| = 3$.

We label the elements of A in increasing order; that is, $A = \{a_1, \ldots, a_n\}$, where $a_i < a_j$ if $i < j$. We also let P be the shortest arithmetic progression containing A, and let d denote the common difference of P. Furthermore, we extend P to the infinite sequence $(p_j)_{j=1}^{\infty}$ given by $p_i = \min(P) + id$, and let $P_j = \{p_1, \ldots, p_j\}$ denote the first j terms of this arithmetic progression.

First, we may assume the removal of any single point does not change the common difference of A - if removing x from A increases the common difference of A, then $A\backslash\{x\}+A\backslash\{x\}$ and $x+A\backslash\{x\}$ are incongruent modulo d, so they are non-intersecting, and the Cauchy-Davenport Theorem gives us the result.

Now we split into cases for the inductive step. We remove the largest element from A, and call the remaining set $B := A \setminus \{a_n\}$. Notice that B still has common difference d, due to the immediately previous argument.

Case 3.1.3.1 (B and its subsets have small diameter)**.**
Assume that $a_j \leqslant p_{2j-2}$ for each $j \in [2, n-1]$. This is equivalent to the fact that P_{2j-2} contains at least j elements of A when $j \in [2, n-1]$, from which we can deduce that P_j contains at least $(j+1)/2$ elements of B for $j \in [3, 2n-3]$.

In particular, the two sets B and $p_j + \min(A) - B$ intersect inside P_j. In otherwords, $\min(A) + p_j \in 2B$ for $j = 3, \ldots, 2n-3$. Since $a_1 = p_1$ and $a_2 = p_2$, we have $\min(A) + P_{2n-3} \subseteq 2B$.

However, as $\operatorname{diam}(A) > 2n-3$, $\min(A) + P_{2n-3}$ does not intersect $\max(A) + A$, and the two sets together contribute at least $(2|A|-3) + |A| = 3|A| - 3$ elements of $2A$, as required.

Case 3.1.3.2 (B has small diameter, but its subsets do not)**.**
Assume that $a_{n-1} \leqslant p_{2n-2}$ but $a_j \geqslant p_{2j-1}$ for some $j \in \{2, \ldots, n-2\}$. Fix i be the largest j for which this is true. Then since we have $p_{2i-1} \leqslant a_i < a_{i+1} < p_{2i+1}$, we know that $a_i = p_{2i-1}$ and $a_{i+1} = p_{2i}$.

Split A into the two sets A_1, A_2 defined by $A_1 = \{a_1, \ldots, a_{i+1}\}$ and $A_2 = \{a_i, \ldots, a_n\}$. Since both sets contain a_i, a_{i+1}, where $a_{i+1} - a_i = d$, notice that A_1 and A_2 have the same common difference d as A.

Now $|A_1| \geqslant 3$ and $\operatorname{diam}(A) = 2i = 2|A_1| - 2$, so applying Lemma 3.1.2 gives

$$|2A_1| \geqslant |A_1| + \operatorname{diam}(A_1) - 1 = 3|A_1| - 3 = 3i.$$

Also, $|A_2| \geqslant 3$, $\min(A_2) = p_{2i-1}$ and $\max(A_2) = \max(A)$, so

$$\operatorname{diam}(A_2) = \operatorname{diam}(A) - (2i-2) \geqslant 2|A| - 2 - (2i+2) = 2|A_2| - 2.$$

The inductive hypothesis then yields $|2A_2| \geqslant 3|A_2| - 3 = 3|A| - 3i$.

All that is left to notice is that $2A_1 \cap 2A_2 = \{2a_i, a_i + a_{i+1}, 2a_{i+1}\}$, so we add the above two lower bounds to get

$$|2A| \geqslant |2A_1| + |2A_2| - 3 \geqslant 3i + (3|A| - 3i) - 3 = 3|A| - 3.$$

Case 3.1.3.3 (B has large diameter)**.**
We may now assume that $a_{n-1} \geqslant p_{2n-1}$. It follows from the inductive hypothesis that $|2B| \geqslant 3|B| - 3 = 3|A| - 6$, so it suffices to show that $|2A \setminus 2B| \geqslant 3$.

The two largest elements of $2B$ are $a_{n-2}+a_{n-1}$ and $2a_{n-1}$, and $2A$ has its three largest elements $a_n+a_{n-2}, a_n+a_{n-1}, 2a_n \in 2A$, the final two of which are larger than either of those elements from $2B$. Thus $2A$ we have found three further elements not included in $2B$ apart from in the case $a_n+a_{n-2}=2a_{n-1}$. So we may assume that the last three elements a_{n-2}, a_{n-1}, a_n of A are in arithmetic progression.

Suppose first that $\{a_{n-2}, a_{n-1}, a_n\}$ has a larger common difference d' than that of A. Take i to be the largest integer such that $\{a_{n-i}, \dots, a_n\}$ is an arithmetic progression. Now if $a_n + a_{n-i-1} \in 2B$ then there are indices $j, k \leqslant i$ such that

$$a_{n-i-1} = a_{n-j} + a_{n-k} - a_n \equiv a_{n-j} \pmod{d'},$$

contradicting the maximality of i. So we may assume that $\{a_{n-2}, a_{n-1}, a_n\}$ has common difference d.

By considering the set $-A$, we determine equally that we may assume $\{a_1, a_2, a_3\}$ is an arithmetic progression with common difference d also.

We now have that

$$a_n \geqslant p_{2n-1}$$

$$a_{n-1} \geqslant p_{2n-3}$$

$$a_{n-2} \geqslant p_{2n-5},$$

so let $i \neq 1$ be the least integer such that $a_i \geqslant p_{2i-1}$; we know that $i \in [4, n-2]$. So define the sets $A_1 = \{a_1, \dots, a_i\}$ and $A_2 = \{a_{i-1}, \dots, a_n\}$, and notice again they both have common difference d.

Now $|A_1| \geqslant 4$ and $\operatorname{diam}(A) \geqslant 2i-1$, the inductive hypothesis tells us

$$|2A_1| \geqslant 3|A_1| - 3 = 3i - 3.$$

Also, $|A_2| \geqslant 4$ and $a_{i-1} \leqslant p_{2(i-1)-2} = p_{2i-4}$, so we have

$$\operatorname{diam}(A_2) \geqslant \operatorname{diam}(A) - (2i-5) \geqslant (2|A|-1) - (2i-5) = 2(|A| - (i-2)) = 2|A_2|$$

and again the induction hypothesis yields

$$|2A_2| \geqslant 3|A_2| - 3 = 3(|A| - i + 2) - 3 = 3|A| - 3i + 3.$$

As $2A_1 \cap 2A_2 = \{2a_i, a_i + a_{i+1}, 2a_{i+1}\}$, we find that

$$|2A| \geqslant |2A_1| + |2A_2| - 3 \geqslant (3i-3) + (3|A| - 3i + 3) - 3 = 3|A| - 3.$$

We may put these two lemmas together straightforwardly to get the required result.

Proof of Theorem 3.1.1. To relate to the previous two lemmas, we restate what we want to prove in our new notation:

Let $A \subseteq \mathbb{Z}$ be a finite set of at least three integers. If $|2A| < 3|A| - 3$ then $\operatorname{diam}(A) \leqslant |2A| - |A| + 1$.

In view of Lemma 3.1.3, it can't be the case that $\operatorname{diam}(A) \geqslant 2|A| - 2$, so we must have $\operatorname{diam}(A) \leqslant 2|A| - 3$. In this case, Lemma 3.1.2 tells us $|2A| \geqslant |A| + \operatorname{diam}(A) - 1$, which rearranges into the claim of the theorem. □

There is an interesting proof of this theorem from a paper of Lev and Smeliansky [LS] which uses the tools of Freiman Isomorphisms we begin to develop in §5.1.

Remark 3.1.4. Now we make a short remark about the inherent imprecision, typical of Freiman-type results, in Theorem 3.1.1. Suppose we start with a set A with doubling $|2A| < 3|A| - 3$. Theorem 3.1.1 then tells us that A has diameter at most $|2A| - |A| + 1 \leqslant 2|A| - 3$.

Now suppose we start with a set A having diameter $\operatorname{diam}(A) = 2|A| - 3$. With this information alone, all we are able to infer is that $\operatorname{diam}(2A) = 4|A| - 7$, and hence only that $|2A| \leqslant 4|A| - 7$, a fair bit larger than what we started with. Thus, in some sense, Theorem 3.1.1 makes quite a loss in its statement. Of course, one could probably argue more precisely and elicit more than just information on the diameter of A.

For example, in the later proof of Freiman's 2.4-Theorem, for example, we not only find that A must have a small diameter, but also that A has a large subset B lying in a much short arithmetic progression, though we do not have the machinery to make this inference yet. One can imagine using this sort of information to make a smaller loss when going from diametrical information to information about doubling, though we will not do this.

4. Small Doubling Constant in Other Groups

4.1. **The Cauchy-Davenport Theorem.** We now consider the famous analogue of Lemma 3.0.4 in residues modulo a prime. By residues modulo a prime p we mean the group $\mathbb{Z}/p\mathbb{Z}$ considered additively.

Theorem 4.1.1 (Cauchy-Davenport)**.** *Let A and B be sets of residues modulo a prime p. Then*

$$|A+B| \geqslant \min\{|A|+|B|-1, p\}.$$

The theorem was originally proven by Cauchy in 1813, and independently rediscovered by Davenport in 1935 [D2]. Davenport's proof [D1] is very short and uses very few ideas, so we go through it to introduce the reader to the idea of transforms, and also the method of proof shall be used later in §4.3. The corresponding inverse theorem, which states the structure of sets achieving this bound, is known as Vosper's theorem, and is discussed later in §4.2.

We shall see that, in application, the Davenport Transform makes short work of the Cauchy-Davenport theorem, as thus was its original purpose. Morally, the Davenport transform is almost exactly the method used by Davenport in proving the above theorem: indeed, our proof of the following proposition is exactly the translation of Davenport's proof into modern notation. Coincidentally, we are not the first to do this, as our exposition matches quite closely, for example, that of [R2].

Proposition 4.1.2 (Davenport Transform)**.** *Let A and B be sets of residues modulo a prime p. Assume that $|B| \geqslant 2$ and $A+B \neq \mathbb{Z}/p\mathbb{Z}$.*

Then there is a proper non-trivial subset $B_2 \subseteq B$ such that $|A+B| \geqslant |A+B_2| + |B| - |B_2|$.

Proof. Since $|B| \geqslant 2$, B generates $\mathbb{Z}/p\mathbb{Z}$, so $A+2B$ contains an element $x \in \mathbb{Z}/p\mathbb{Z}$, say, that is not in $A+B$. Let B_1 be the largest subset of B such that $x - B_1 \subseteq A+B$, and then let $B_2 = B \setminus B_1$. Obviously, both B_1 and B_2 are non-empty (if this is not clear, write x as a sum of elements from A and B).

Note that $(x - B_1) - B_2$ does not contain any elements of A — otherwise, there exist $a \in A$ and $b, b_2 \in B_2$ such that $x - b - b_2 = a$; that is, $x - b_2 = a + b$, patently false — hence $A + B_2$ and $x - B_1$ are distinct sets. This means

$$|A+B_2| \leqslant |A+B| - |x-B_1| = |A+B| - |B_1| = |A+B| - |B| + |B_2|. \quad \square$$

Here, the start of the proof is slightly modified from Davenport's proof; Davenport appeals to the inductive hypothesis to find the element x. In any case, the main part of the work is done, so the proof of Theorem 4.1.1 becomes very short.

Davenport's Proof of Theorem 4.1.1. We work by induction on $|B|$. The theorem is trivial for $|B| = 1$ and $|B| = 2$, so we make the inductive assumption $|B| \geqslant 3$ and assume the theorem holds for all smaller sets. We also assume that $|A+B| < p$, so we need only prove that $|A+B| \geqslant |A| + |B| - 1$.

With immediate respect to Proposition 4.1.2, there is a Davenport transform B_2, say, of B, such that $|A+B| \geqslant |A+B_2| + |B| - |B_2|$. However, by the inductive hypothesis,

$$|A+B_2| \geqslant |A| + |B_2| - 1,$$

so that

$$|A+B| \geqslant |A| + |B| + |B_2| - 1 - |B_2| \qquad \square$$

There are other short proofs of the Cauchy-Davenport theorem that utilise transforms. Dyson's e-transform is seemingly the most well-known and oft-used transform, and Chowla's proof of the Cauchy-Davenport theorem utilising the e-transform can be found in [N].

4.2. **Vosper's Theorem.** Vosper's theorem was the first non-trivial inverse theorem in residues modulo a prime. The most interesting part of the theorem, as far as the author is concerned, is the following.

Theorem 4.2.1 (Vosper's Theorem)**.** *Let A and B be sets of residues modulo a prime p. Assume that $|A| \geqslant 2, |B| \geqslant 2$ and $|A+B| < p-1$. Then $|A+B| = |A|+|B|-1$ if and only if both A and B are arithmetic progressions with the same common difference.*

Vosper proves this using the Davenport transform in [V1]. Much of that paper concerns the border cases where, for example $|A| = 1$ or $|B| = 1$ or both $|A|$ and $|B|$ are large. For the most part, these situations become a case-by-case check that the theorem doesn't fail in such situations, and do not make particularly enthralling read. In any case, shortly thereafter, Vosper offered a simpler proof by utilising Dyson's e-transform [V2].

Assuming we do not utilise Kneser's theorem, which renders Vosper's theorem a mere footnote, then as far as the author is aware the shortest known proof of Vosper's Theorem utilises the Davenport-transform, and is due to Rodseth [R2]. We present this proof here to showcase it's elegance.

Proof of Theorem 4.2.1. We look for a contradiction, so let A, B be a pair of sets satisfying $|A + B| = |A| + |B| - 1$ with A not an arithmetic progression, and $|B|$ minimal. We may assume without loss of generality that $0 \in B$.

As in the construction of the Davenport transform, we let $X = (A + 2B) \setminus (A + B)$ and for $x \in X$ we let B_x denote the corresponding Davenport transform of B. This means that

$$\begin{aligned} |A + B_x| &\leqslant |A + B| - |B| + |B_x| \\ &= |A| + |B_x| - 1 \end{aligned}$$

so that $|B_x| = 1$, by the minimality of $|B|$. We thus know that $|B_x|$ is a singleton for every $x \in X$. Since we have assumed $0 \in B$, if we look at the definition of the Davenport transform it turns out firstly that $B_x = \{0\}$ and secondly that $X - (B \setminus \{0\}) \subseteq A + B$.

So let $B' = B \setminus \{0\}$. Then $A \cup (X - B') \subseteq A + B$ and $A \cap (X - B') = \emptyset$, so of course $|A+B| \geqslant |A|+|X-B'|$. Thus, using the Cauchy-Davenport theorem in the second line,

$$\begin{aligned} |A| + |B| - 1 &\geqslant |A| + |X - B'| \\ &\geqslant |A| + |X| + |B'| - 1 \\ &\geqslant |A| + |X| + |B| - 2. \end{aligned}$$

So $|X| \leqslant 1$ and X, too, is a singleton – that is, $A + 2B$ has exactly one more element than $A + B$ – and since $|A + B| < p - 1$ we have $|A + 2B| < p$. So, applying the Cauchy-Davenport theorem again, we

find

$$\begin{aligned} 1 &= |A+2B| - |A+B| \\ &\geqslant |A+B| + |B| - 1 - |A+B| \\ &= |B| - 1. \end{aligned}$$

So $|B| = 2$. Thus B is an arithmetic progression. It is now an easy observation that if A were not an arithmetic progression with the same common difference, then $|A+B| \geqslant |A|+2 = |A|+|B|$, so A is an arithmetic progression after all. This contradiction completes the theorem. □

Remark 4.2.2. The proof is very short and seemingly simple, but within it contains several subtle ideas. The most powerful idea, of course, is that of the Davenport transform, but it is the following more subtle idea that finishes off the proof: that B being an arithmetic progression implies that A is an arithmetic progression. It is a simple observation that we shall mention again in §4.3 where we discuss a theorem that goes 'one step beyond' Vosper's Theorem.

4.3. **A Theorem of Hamidoune and Rodseth.** In the spirit of going beyond the Cauchy-Davenport Theorem in the direction of a result resembling Freiman's $3k-3$ Theorem, we now discuss the following result of Hamidoune and Rodseth. To state the theorem, we need to introduce the notion of a *puncture*. Let A be an arithmetic progression, and $a \in A$. We say that the set $B = A \setminus \{a\}$ is a *punctured arithmetic progression*. Equivalently, a set B is a punctured arithmetic progression if $\mathrm{diam}(B) \leqslant |B|+1$.

A punctured progression is referred to as an *almost-progression* in [HR]. Later, in §8.5, we shall consider cosets with points missing which we refer to as *punctured cosets*, so we use our term *puncture* to stay consistent.

Theorem 4.3.1. *Let A and B each be sets of at least three residues modulo a prime p. If*

$$7 \leqslant |A+B| \leqslant |A| + |B| \leqslant (p-4)$$

then A and B are punctured progressions with the same common difference.

The theorem is resolved in the most part in a single theorem, whose proof is almost identical to the proof of Vosper's Theorem showcased above.

Theorem 4.3.2. *Let A and B each be sets of at least two residues modulo a prime p. If*

$$|A+B| \leqslant |A| + |B| \leqslant p-4$$

then A — and hence B, too — is a union of two arithmetic progressions.

The rest of the paper becomes a systematic breakdown of the the possible consequences of this theorem with a careful refinement of the idea mentioned in Remark 4.2.2. From the above theorem, if we make the assumption that $|B| \geqslant 3$, for instance, then B must contain an arithmetic progression of length two. Thus, as at the end of the proof of Vosper's Theorem, if we consider adding just this fragment of an

arithmetic progression to A then if A itself is not an arithmetic progression with the same common difference then we are guaranteed to gain two extra elements to the sumset $A + B$.

So, if we call an arithmetic progression of length two a *fragment*, the start of the paper [HR] makes a few simple observations of how punctured arithmetic progressions and unions of two arithmetic progressions respond when they are added to a fragment with the same common difference. It then turns out that one can make similar inferences to the above by carefully finding these fragments in the sets A, B and $A + B$ and pitting them against each other. It turns out there are quite a few cases to check when using this argument, and the paper [HR] clearly shows hard graft in making sure all the remaining cases are mopped up. Because of its case-by-case nature, the proof is long but straightforward.

So, this simple idea, when worked meticulously, can take us one step beyond Vosper's Theorem.

Remark 4.3.3. I recently became aware that, in the paper [R2], Rodseth states that in the paper of Hamidoune, Serra and Zemor [HSZ] there is a result of that goes one step further than Theorem 4.3.1 — presumably, a characterisation of sets A, B such that $|A+B| \leqslant |A|+|B|+1$, provided $|A|$ and $|B|$ are suitably restricted.

The mentioned paper makes heavy use of so-called isoperimetric tools about which the author is not familiar, and hence it is not clear where at all this improvement is made. Indeed, there appears to be no mention of the result in that paper; rather, the only results mentioned in this vain are Vosper's Theorem and Theorem 4.3.1.

In any case, this seems to be the full extent of results that characterise sets completely when their doubling is almost exactly 2. Known further characterisations are much less precise, as we shall now go on to discuss. In the coming section, we begin to discuss a more statistical approach coming primarily from Fourier analysis to gain similar results.

5. Fourier Analytic Methods

This section shall see us attempt a different method in proving inverse theorems in sets of residues modulo a prime, utilising the ideas of Fourier analysis, Freiman isomorphisms and rectification.

5.1. **Freiman homomorphisms.** Now, we consider an idea that has proved powerful in the study of small-doubling questions in the integers – namely, the use of Freiman homomorphisms to transfer additive problems between different additive groups. Addition may be studied more easily in one group than another, so we formalise a way in which we can interchange between the two.

Definition 5.1.1 (Freiman homomorphisms, Freiman isomorphisms)**.** Let A be a finite set of elements from the group G, and B a finite set of elements from the group H, where G and H are both abelian groups written additively, and let $\varphi : A \to B$ be a function. If it is the case that, whenever $a, b, c, d \in A$ satisfy $a + b = c + d$, we also have that $\varphi(a) + \varphi(b) = \varphi(c) + \varphi(d)$ then we call φ a Freiman homomorphism from A to B. If φ^{-1} exists and is also a Freiman homomorphism from B to A, we say that φ is a Freiman isomorphism, and also that A and B are Freiman isomorphic.

It is sometimes not obvious how to approach a question of small doubling in sets of residues modulo a prime, whereas answering a similar question in the integers is more obvious, as for example we are able to use the ordering of the integers in our arguments. Conversely, the integers modulo a prime make up a field, so if we can change from a question about integers to a question about residues then we might be

able to use the structure of the field to make inferences that would not otherwise be obvious.

We shall later see several examples of utilising results in the integers to prove results about residues modulo a prime, so now we give a result of Erdös [E], of the converse situation for illustrative purposes. We present the result as found [TV].

A sum-free set is one such that $A+A$ does not contain any elements of A. Notice that a sum-free set cannot contain zero.

Lemma 5.1.2. *Let A be a finite set of integers. Then A contains a sum-free subset of size larger than $|A|/3$.*

Proof. Let p be a large prime and such that $p = 3k + 2$ for some integer k and so that $-p/4 < \min(A) < \max(A) < p/4$. Let $\varphi : A \to \mathbb{Z}/p\mathbb{Z}$ be the function that maps an integer $a \in A$ to the corresponding residue $a \in \mathbb{Z}/p\mathbb{Z}$.

Notice that $\varphi(A) \subseteq (-p/4, p/4)$, so that $2\varphi(A) \subseteq (-p/2, p/2)$ — that is, addition of two elements in $\varphi(A)$ cannot 'wrap around, so φ is clearly a Freiman isomorphism between A and $\varphi(A)$.

Now let $x \in \mathbb{Z}/p\mathbb{Z}$ be an arbitrary non-zero residue, and consider the set

$$A_x := x\varphi(A) \cap [k+1, 2k+1].$$

A_x is clearly a sum-free subset of $\mathbb{Z}/p\mathbb{Z}$, and hence $x^{-1}A_x$ is a sum-free subset of $\varphi(A)$. Furthermore, $\varphi^{-1}(x^{-1}A_x)$ is a sum-free subset of A. Thus all we need to do is find some x such that $|A_x| > |A|/3$.

To do this, we show that the expected size of $|A_x|$ for a random $x \in \mathbb{Z}/p\mathbb{Z}$ is at least $|A|/3$. So, notice that

$$\mathop{\mathbb{E}}_{x\in\mathbb{Z}/p\mathbb{Z}} A_x = \sum_{a\in A} \mathbb{P}_{x\in\mathbb{Z}/p\mathbb{Z}}(x\varphi(a) \in [k+1, 2k+1]).$$

However, since x was non-zero, the random variable $x\varphi(a)$ is uniformly distributed in the non-zero elements of $\mathbb{Z}/p\mathbb{Z}$, and $\mathbb{P}(x\varphi(a) \in [k+1, 2k+1]) > 1/3$ for all $a \in A$. This means that $\mathbb{E}_{x\in\mathbb{Z}/p\mathbb{Z}} |A_x| > |A|/3$. Consequently there must be some non-zero $x \in \mathbb{Z}/p\mathbb{Z}$ that achieves this bound, as required. □

As well as demonstrating the use of a Freiman homomorphism, this also demonstrates the principle of probabilistic arguments. Note that the existence of a specific x is not given, but we calculate a probability to show that a random choice of x has positive probability of achieving the required bound. This is simply an application of the pigeonhole principle; there must be some choice of x achieving at least the expected value.

This is the simplest version of a probabilistic argument, but we shall not discuss these ideas any further. We shall use this same argument repeatedly in §7. The interested reader can find an extremely thorough discussion of this and other types of probabilistic argument in [TV].

5.2. **Rectification and Small Doubling in Residues.** Addition in the integers and addition in residues modulo a prime take on a very similar form to one other, so in this section we start to consider small-doubling problems in sets of residues by using Freiman isomorphisms to transfer the problem into questions about sets of integers. This process is called rectification.

Definition 5.2.1. We say that a set of residues A modulo a prime is *rectifiable* if there is a function $\varphi : A \to \mathbb{Z}$ such that φ is a Freiman isomorphism between A and $\varphi(A)$.

The point of rectification is that the integers, for the most part, are easier to deal with. Indeed, we have already seen proven some results in the integers that were trivial to prove whereas the analogous result in residues was not so straightforward. For example, Lemma 3.0.1 required almost no proof, but the Cauchy-Davenport Theorem required a new idea that was not so intuitive.

There are a few papers in the literature concerning rectification of sets with small doubling. Indeed, in his endeavour to prove Freiman's theorem, Freiman [F] proves a result of the following type.

Theorem 5.2.2. *Let $c > 1$ be a real number, and p a prime. Then there is a real number $\alpha \in 0, 1$ depending only on c such that if a set $A \subseteq \mathbb{Z}/p\mathbb{Z}$ satisfies both $|A| < \alpha p$ and $|2A| < c|A|$ then A is rectifiable.*

Remark 5.2.3. The result says that small sets with small doubling are rectifiable. While useful in principle, the number α proved by Freiman decreases exponentially with c. This was improved slightly by Bilu, Lev and Ruzsa in [BLR]. As far as the author is aware, the best known result of this type is currently due to Green and Ruzsa [GR], where it is shown that we may take $\alpha \geqslant (32c)^{-12c^2}$.

For our applications that follow, these bounds are too weak. Fortunately, as is often the case when considering extremely small c, there are arguments that allow us to take α much larger. The most useful for us, and the best currently known, is a result of Lev [L1]. Rather than showing that a set with small sumset is rectifiable, it shows that a set

with a large Fourier coefficient has a large subset which is rectifiable. While, at the moment, this may seem unrelated, we shall see that small doubling and large Fourier coefficients go hand-in-hand. This will be our first foray into Fourier analysis.

Remark 5.2.4. Firstly, we reintroduce our earlier definition of diameter in a setting where it is wholly more relevant. For a set A of residues modulo a prime p we define, as before, the diameter $\mathrm{diam}(A)$ of A to be the length of the shortest arithmetic progression containing A.

Secondly, we mention a simple case in which rectification is an easy conclusion. If a set of residues A modulo a prime p has diameter at most $(p+1)/2$ then A is clearly rectifiable. Indeed, first notice that an affine transformation will send A into $[0,(p-1)/2]$. From there, map the elements of A to the smallest non-negative integer in the equivalence class. The composition of these two Freiman isomorphisms is trivially also a Freiman isomorphism.

Remark 5.2.5. Now we discuss the innevitable appearance of Fourier analaysis here, thanks to the obvious isomorphism between the additive abelian group $\mathbb{Z}/p\mathbb{Z}$ and the multiplicative abelian group U_p, the pth roots of unity, as mentioned in §2.2. Specifically, the isomorphism $\gamma : \mathbb{Z}/p\mathbb{Z} \to U_p$ is given by

$$\gamma(a) = \exp(2i\pi a/p)$$

for all $a \in \mathbb{Z}/p\mathbb{Z}$ — it's really an abuse of notation to put a in the exponent here, but notice that any representative of $a \in \mathbb{Z}/p\mathbb{Z}$ will give the same value for γ, so we shall continue to do this and assume that whenever we put an element $a \in \mathbb{Z}/p\mathbb{Z}$ in an exponent we mean

the smallest positive integer in the equivalent class. See also Remark 5.2.6.

This is an incredibly useful visualisation for subsets of $\mathbb{Z}/p\mathbb{Z}$; we can picture them as points on the circle which, when added, merely add arguments. Now, consider a set $A \subseteq [0, (p-1)/2]$. Then, in this visualisation, A lies in an closed arc of length $\pi - (2\pi)/p$ on the unit circle in the complex plane. So, we have a nice way of picturing sets of residues that are obviously rectifiable.

Now notice that, in such a case, the sum $\sum_{a\in A} \gamma(a)$ as a vector in the complex plane will point in some direction along this same arc on which $\gamma(A)$ lies. Also, the sum will be relatively large compared to what you would expect if A was an arbitrary set, as there is no contribution from the elements on the opposing arc to the sum. So, in some sense, this sum measures how heavily weighted the set A is to one side of the circle in the complex plane.

The final thing to notice is that the sum $\sum_{a\in A} \gamma(a)$ is

$$\sum_{a\in A} \gamma(a) = p \times \mathop{\mathbb{E}}_{a\in A} \gamma(a) = p \times \widehat{1_A}(\gamma).$$

Thus the size of Fourier coefficient $\widehat{1_A}(\gamma)$ measures how heavily weighted the set $\gamma(A)$ is to one side of the circle in the complex plane. Returning to the start of this discussion, we see that $|\widehat{1_A}(\gamma)|$ being large implies, at least heuristically, that more of A lies in an arithmetic progression of length $(p+1)/2$ with common difference 1. Similarly, if the Fourier transform $\widehat{1_A}(\gamma')$ of any other character $\gamma' \in \widehat{G}$ is large, then the same argument tells us that A has a large subset lying in an arithmetic progression of length $(p+1)/2$ with some other common difference, because

the other characters on $\mathbb{Z}/p\mathbb{Z}$ take the form

$$a \mapsto \exp(2i\pi da/p)$$

for some $d \in \mathbb{Z}/p\mathbb{Z}$. In summary, heuristically speaking, if $\max_{\gamma \neq \gamma_0} |\widehat{1_A}(\gamma)|$ is large, then it suggests that A has a large rectifiable subset. We shall present a theorem of Lev that indeed shows this is the case.

Remark 5.2.6. As we mentioned in Remark 5.2.4, we will abuse notation slightly when considering residues appearing in exponents. We shall do so without further comment in other circumstances too. For example, later in the proof of Lemma 5.4.4 we shall even go so far as to use inequalities with residues. It will be obvious from the context that we mean the inequalities hold for the smallest integer representatives of the residues in question and shall not worry about the consequences of doing so. There are other instances too that the reader may spot.

This is not done with the intention of being innaccurate. Rather, the added detail in making the arguments more precise only detract from the discussion and serve to hide the techniques. In all cases, writing out the offending arguments with the added details should persuade the reader of their obviousness.

Now, we return to the result of Lev [L1]. Though in spirit our proof below is the same result of Lev, a fleshed out version of the below proof is best found in a paper of Green [G] which utilises a near-identical argument.

Proposition 5.2.7. *Let A be a set of residues modulo a prime p and let*

$$\gamma(a) = \exp(2i\pi a/p).$$

Then $|\widehat{1_A}(\gamma)| \leqslant |\widehat{1_B}(\gamma)|$*, where* B *is an arithmetic progression of length* $|A|$ *and common difference* 1.

Proof. Let $A \subseteq \mathbb{Z}/p\mathbb{Z}$ be a set achieving the maximum possible value for $|\widehat{1_A}(\gamma)|$. If $A = \mathbb{Z}/p\mathbb{Z}$, the result is obviously true, so we insist $|A| < p$. Notice that this rules out $|\widehat{1_A}(\gamma)| = 0$.

Now picture the points of $\gamma(A)$ lying around the circle, and their sum

$$S := \sum_{a \in A} \exp(i\pi a/p)$$

pointing in the the direction of the unit vector $\exp(2\pi\theta/p)$, say. If any element of $\gamma(A)$ was replaced by an element of $\gamma[(\mathbb{Z}/p\mathbb{Z}) \setminus A]$ that was strictly closer to the vector $\exp(2\pi\theta/p)$ then, by considering the appropriate parallelogram, the resulting set would have a larger corresponding sum $|S|$. Thus there can be no points in $\gamma(\mathbb{Z}/p\mathbb{Z})$ closer to $\exp(2\pi\theta/p)$ that are not already elements of $\gamma(A)$, so $\gamma(A)$ is clustered as closely around the vector $\exp(2\pi\theta/p)$ as possible.

Consequently, the points of $\gamma(A)$ are equally spaced around the vector $\exp(2\pi\theta/p)$, and hence A must be an arithmetic progression with difference 1. □

Remark 5.2.8. For the rest of this paper, we let $\gamma_0 \in \widehat{\mathbb{Z}/p\mathbb{Z}}$ denote the trivial character given by $\gamma_0(x) = 1$ for all $x \in \mathbb{Z}/p\mathbb{Z}$. One may check that the other $p-1$ non-trivial characters of $\mathbb{Z}/p\mathbb{Z}$ take the form

$$a \mapsto \exp(2i\pi da/p)$$

for some $d \in \mathbb{Z}/p\mathbb{Z}$.

Theorem 5.2.9 (Lev)**.** *Let* A *be a set of residues modulo an odd prime* p*. Let* B *be the largest subset of* A *with diameter* $\operatorname{diam}(B) \leqslant (p+1)/2$*.*

Then

$$\max_{\substack{\gamma\in\widehat{\mathbb{Z}/p\mathbb{Z}}\\ \gamma\neq\gamma_0}} |\widehat{1_A}(\gamma)| \leqslant \frac{\sin\left[\left(|B|-\frac{|A|}{2}\right)\frac{2\pi}{p}\right]}{p\sin\left(\frac{\pi}{p}\right)}.$$

Proof. Let A be an arithmetic progression. By Proposition 5.2.7 if we can prove the result for this set A, it follow for all sets A of the same size.

Without loss of generality, assume that A has common difference 1 and retain the notations of γ and S from the proof of Proposition 5.2.7. There's no loss of generality here because the statement holds for any dilate of A, and the other $p-1$ non-trivial characters are just the character γ calculated on the dilates of A, as we see from Remark 5.2.8 — this excludes the trivial character γ_0.

It is a simple observation that $|\widehat{1_A}(\lambda)|$ achieves its maximum at $\lambda=\gamma$, since for instance the corresponding sum $|S|$ is the highest attainable among all sums of $|A|$ elements of $\gamma(\mathbb{Z}/p\mathbb{Z})$. Moreover, as S is a geometric progression, this maximum is easy to calculate as

$$|\widehat{1_A}(\gamma)| = \frac{1}{p}\times|S| = \frac{1}{p}\times\frac{\sin\left(|A|\frac{\pi}{p}\right)}{\sin\left(\frac{\pi}{p}\right)}.$$

This bound will give us the theorem, provided we consider two distinct cases. The first is that $|A|\leqslant(p+1)/2$, in which case $|B|=|A|$ and

$$\frac{|A|}{2} = |B| - \frac{|A|}{2}$$

which is what we wanted. The second case is that $|A| > (p+1)/2$, in which case $|B| = (p+1)/2$ and

$$\begin{aligned}\sin\left(|A|\frac{\pi}{p}\right) &= \sin\left(\pi - |A|\frac{\pi}{p}\right)\\ &= \sin\left[\left(\frac{p}{2} - \frac{|A|}{2}\right)\frac{2\pi}{p}\right]\\ &\leqslant \sin\left[\left(\frac{p+1}{2} - \frac{|A|}{2}\right)\frac{2\pi}{p}\right]\\ &= \sin\left[\left(|B| - \frac{|A|}{2}\right)\frac{2\pi}{p}\right].\end{aligned}$$

The inequality holds because we can tell from the first line, for instance, that the operand is in $[0, \pi/2)$ — though we need to be careful in the case $|A| = (p+3)/2$; that is where equality can occur. So, in either case, we are done. □

Corollary 5.2.10. *Let A be a set of residues modulo an odd prime p. Let B be the largest rectifiable subset of A. Then*

$$|B| \geqslant \frac{|A|}{2} + \frac{p}{2\pi}\arcsin\left[p\sin\left(\frac{\pi}{p}\right)\max_{\substack{\gamma\in\widehat{\mathbb{Z}/p\mathbb{Z}}\\ \gamma\neq\gamma_0}}|\widehat{1_A}(\gamma)|\right]$$

Proof. Note that, from Remark 5.2.4, a set with diameter at most $(p+1)/2$ is rectifiable, so we may use Theorem 5.2.9; the bound here is the inequality of that Theorem rearranged to make $|B|$ the subject. □

5.3. **Fourier Analysis, Convolution and Small Doubling in Residues.** This section starts to tie together a few ideas that we have discussed so far. It shows how the upcoming theorems of Freiman use a beautiful convergence of several concepts.

In view of Corollary 5.2.10, if we want to find rectifiable subsets of sets with small doubling then we need only show that sets with small doubling have a large Fourier coefficient, which won't take us very long at all. So now we briefly consider Fourier analysis on sets of residues. In particular, we shall prove the following.

Theorem 5.3.1. *Let A be a subset of the finite abelian group G. If $|2A| = c|A|$ and $|A| = \alpha p$ then*

$$\max_{\substack{\gamma \in \widehat{\mathbb{Z}/p\mathbb{Z}} \\ \gamma \neq \gamma_0}} |\widehat{1_A}(\gamma)| \geqslant \alpha \sqrt{\frac{1-c\alpha}{c(1-\alpha)}}$$

Proof. The proof of this will stem from considering the quantity

$$\mathop{\mathbb{E}}_{x \in G} (1_A * 1_A)(x).$$

If recall from §2.2 how convolution works, we know that for each $x \in G$, $p \times 1_A * 1_A(x)$ is the number of ways that x can be written as a sum of elements $a + b$ where $a, b \in A$. Thus the above sum simply counts all the pairs $(a, b) \in A$ and divides it by $|G|^2$. That is

$$\alpha^2 = \mathop{\mathbb{E}}_{x \in G} (1_A * 1_A)(x)$$

However, since the $1_A * 1_A$ is supported on 1_{2A}, we may write

$$\alpha^2 = \mathop{\mathbb{E}}_{x \in G} (1_A * 1_A)(x)\overline{1_{2A}(x)}.$$

Now we reconsider §2.5, and Parseval's identity in particular. It tells us that

$$\begin{aligned}\alpha^2 &= \mathop{\mathbb{E}}_{x\in G}(1_A * 1_A)(x)\overline{1_{2A}(x)}\\ &= \sum_{\gamma\in\widehat{G}}\widehat{1_A * 1_A}(\gamma)\overline{\widehat{1_{2A}}(\gamma)}\\ &= \sum_{\gamma\in\widehat{G}}\widehat{1_A}(\gamma)^2\overline{\widehat{1_{2A}}(\gamma)},\end{aligned}$$

where in the last equality we've used the fact that convolution of functions turns into multiplication of functions under the Fourier transform. Now, most of the work is done. Firstly, we take out the trivial character γ_0 since we can calculate the contribution from that term exactly, namely

$$\begin{aligned}\alpha^2 &= \sum_{\gamma\in\widehat{G}}\widehat{1_A}(\gamma)^2\overline{\widehat{1_{2A}}(\gamma)}\\ &= \alpha^2\times c\alpha + \sum_{\gamma\neq\gamma_0}\widehat{1_A}(\gamma)^2\overline{\widehat{1_{2A}}(\gamma)}\\ &\leqslant c\alpha^3 + \max_{\gamma\neq\gamma_0}|\widehat{1_A}(\gamma)|\sum_{\gamma\neq\gamma_0}|\widehat{1_A}(\gamma)||\overline{\widehat{1_{2A}}(\gamma)}|.\end{aligned}$$

Now we can utilise the Cauchy-Schwarz inequality, in the form

$$\begin{aligned}\alpha^2 &\leqslant c\alpha^3 + \max_{\gamma\neq\gamma_0}|\widehat{1_A}(\gamma)|\sum_{\gamma\neq\gamma_0}|\widehat{1_A}(\gamma)||\overline{\widehat{1_{2A}}(\gamma)}|\\ &\leqslant c\alpha^3 + \max_{\gamma\neq\gamma_0}|\widehat{1_A}(\gamma)|\left(\sum_{\gamma\neq\gamma_0}|\widehat{1_A}(\gamma)|^2\right)^{1/2}\left(\sum_{\gamma\neq\gamma_0}|\widehat{1_{2A}}(\gamma)|^2\right)^{1/2}.\end{aligned}$$

Now we use Parseval's identity one more time to work out these bracketed quantities. From our discussion in this proof already, it should be

clear that

$$\sum_{\gamma \neq \gamma_0} |\widehat{1_A}(\gamma)|^2 = \alpha - \alpha^2$$

and

$$\sum_{\gamma \neq \gamma_0} |\widehat{1_{2A}}(\gamma)|^2 = c\alpha - (c\alpha)^2,$$

so returning to our previous inequality we thus far have

$$\alpha^2 \leqslant c\alpha^3 + \max_{\gamma \neq \gamma_0} |\widehat{1_A}(\gamma)| \sqrt{\alpha(1-\alpha)} \sqrt{c\alpha(1-c\alpha)}.$$

The theorem is now proved by rearranging this inequality to make $\max_{\gamma \neq \gamma_0} |\widehat{1_A}(\gamma)|$ the subject. □

Remark 5.3.2. Each of the steps in the proof above are standard fare, and the proof is presented somewhat longer than is necessary. Utilising Parseval's identity and the Cauchy-Schwarz inequality are so frequent in Fourier analytic arguments that they are scarcely remarked on throughout the literature, and we shall scarcely remark on them in the future. With this in mind, we write here how the proof of Theorem 5.3.1 can be reduced to a simple few lines of equations when not all the steps are painstakingly explained.

Concise Proof of Theorem 5.3.1 We have

$$\begin{aligned}
\alpha^2 &= \sum_{\gamma \in \widehat{G}} \widehat{1_A}(\gamma)^2 \overline{\widehat{1_{2A}}(\gamma)} \\
&= c\alpha^3 + \sum_{\gamma \neq \gamma_0} \widehat{1_A}(\gamma)^2 \overline{\widehat{1_{2A}}(\gamma)} \\
&\leqslant c\alpha^3 + \max_{\gamma \neq \gamma_0} |\widehat{1_A}(\gamma)| \left(\sum_{\gamma \neq \gamma_0} |\widehat{1_A}(\gamma)|^2 \right)^{1/2} \left(\sum_{\gamma \neq \gamma_0} |\widehat{1_{2A}}(\gamma)|^2 \right)^{1/2} \\
&= c\alpha^3 + \max_{\gamma \neq \gamma_0} |\widehat{1_A}(\gamma)| \sqrt{\alpha(1-\alpha)} \sqrt{c\alpha(1-c\alpha)}
\end{aligned}$$ □

Remark 5.3.3. Note that we have stated Theorem 5.3.1 for finite abelian groups G rather than just $\mathbb{Z}/p\mathbb{Z}$. This is because the proof utilises no information whatsoever about $\mathbb{Z}/p\mathbb{Z}$, and also allows us to use the result straight from here when we require another large Fourier coefficient estimate in §6

Remark 5.3.4. Freiman proved a result similar to, but weaker than, Theorem 5.3.1 [F]. His proof is identical except that he does not take into account the contribution from the trivial character.

5.4. **Freiman's** 2.4 **Theorem, and Improvements.** Now we go on to discuss and prove the following well-known result of Freiman.

Theorem 5.4.1 (Freiman's 2.4-Theorem)**.** *Let A be a set of residues modulo a prime p. If $|A| \leqslant p/35$ and $|2A| < 2.4|A|$ then* $\operatorname{diam}(A) \leqslant |2A| - |A| + 1$.

The conditions on the result are quite restrictive, and as we shall show these conditions may be relaxed somewhat, though not perhaps as much as we would like.

Remark 5.4.2. Notice that this result is equivalent to a rectification result. Notice that since

$$|2A| - |A| + 1 < 2.4(p/35) + 1 < p/2$$

the result tells us that A is rectifiable. Conversely, if A was rectifiable, then the result follows simply by considering Theorem 3.1.1. So our discussion hasn't really changed from rectification at all, and thus the principle steps in the proof of Freiman's 2.4-Theorem are as follows.

(i) Show the small doubling infers a large Fourier coefficient;

(ii) use the large Fourier coefficient to find a large rectifiable subset;

(iii) use this set to show the whole set is rectifiable.

Each of the above steps in the proof of Theorem 5.4.1 that we present is stronger than the corresponding steps in Freiman's proof [F], though we shall explain what Freiman did differently, as for instance we have already done in Remark 5.3.4. This will make our proof a clear exposition of Rodseth's improvement to the Freiman 2.4-Theorem, though all the steps were done independently and previous to the author's discovery of that paper. We shall talk briefly about this later.

For the second step, we shall use the result of Lev we proved earlier (Corollary 5.2.10), whereas Freiman [F2] utilised the following weaker result.

Theorem 5.4.3 (Freiman)**.** *Let A be a set of residues modulo an odd prime p. Let B be the largest rectifiable subset of A. Then*

$$|B| \geqslant \frac{|A|}{2} + \frac{p}{2} \max_{\gamma \neq \gamma_0} |\widehat{1_A}(\gamma)|.$$

This can easily be deduced from Corollary 5.2.10. Indeed, if one considers the statement there and writes it in terms of the function $\sin(x)/x$, then using the fact that $|\sin(x)/x| \leqslant 1$ for all $x \in \mathbb{R}$ will give the result of Freiman. Freiman's proof, on the other hand, is intricate, complicated and hardly transparent, as he splits up the circle into various components and recombines them in various ways to get the result [F]. The reader interested in these rectification arguments should check out should check Lev's two papers [L1] and [L2] where he proves further generalisations.

Thus, to finish our proof of Freiman's 2.4-Theorem, we need only do the final step. That is, we need to show that a set that has a

large rectifiable subset is itself rectifiable. So, this is what we do next, via a simple combinatorial lemma. The first part of the combinatorial lemma is due to Freiman. The second part is an improvement due to an argument of the author.

Lemma 5.4.4 (The Packing Lemma)**.** *Let $A \subseteq \mathbb{Z}/p\mathbb{Z}$ be a set of residues modulo a prime p. Suppose that $|2A| < c|A| - 3$ and suppose that A has a subset B of size at least $(c/3)|A|$ and diameter L. If $L < p/3$ then $\mathrm{diam}(A) < 3L$ (weak form).*

Furthermore, if $L < p/4$ then $\mathrm{diam}(A) < 2L$ (strong form).

Proof. It may be illustrative to the reader to envisage these sets on the circle via the Freiman isomorphism described in Remark 5.2.5, as the following claims then become immediate.

We may dilate and translate A to assume that, without loss of generality, $B \subseteq [0, L-1]$ and $0 \in B$. Let $x \in A \setminus B$. Firstly, note that $x \in [2L-1, p-L]$ else then the set $x + B$ is distinct from the set $2B$, so the Cauchy-Davenport Theorem tells us

$$|2A| \geqslant |2B| + |B| \geqslant 3|B| - 1 > c|A| - 1.$$

Hence $x \in [-(L-1), 2L-2]$; that is, $A \subseteq [-(L-1), 2L-2]$, which proves the weak form.

To deduce the second part, let $x, -y$ be the 'outermost elements' of A, meaning that $\mathrm{diam}(A) = x + y + 1$. Consider the two sets

$$C = -y + B \supseteq (-y + B) \cap [-y, -1] = -y + B \cap [0, y-1]$$

$$D = x + B \supseteq (x + B) \cap [2L-1, L-1-x] = x + B \cap [2L-1-x, L-1].$$

If $L < p/4$ — coming from the worst case scenario of $x = 2(L-1)$ and $y = -(L-1)$ — then the sets C, D do not intersect. Furthermore, they are both contained in $2A$ and are disjoint from $2B$. By the latter equalities in each of the above lines, these two new sets contribute at least

$$|B \cap ([0, y-1] \cup [2L-1-x, L-1])|$$

new elements to the sumset $2A$. So if the two intervals $[0, y-1]$ and $[2L-1-x, L-1]$ cover B then this is a contribution of at least $|B|$ new elements to the sumset $2A$ and we get a contradiction by the same argument as in the first part, indicating that $y - 1 < 2L - 1 - x$ or, in particular,

$$x + y + 1 < 2L.$$

Recalling that $x + y + 1 = \text{diam}(A)$, this proves the strong form. □

Now all we have to do to finish the Freiman 2.4-Theorem is combine Theorem 5.3.1, Corollary 5.2.10 and Lemma 5.4.4. Note that, as already mentioned, this will actually give us an improvement on the Freiman 2.4-Theorem as each of the three results are stronger than the corresponding component of Freiman's original proof. We shall put the result this into context after its proof.

Theorem 5.4.5. *Let $\alpha, c \in \mathbb{R}$ be such that $c \in [2,3]$, $\alpha \in [0, 3/8c]$ and*

$$\frac{1}{2\pi} \arcsin\left(\pi\alpha\sqrt{\frac{1-c\alpha}{c(1-\alpha)}}\frac{\sin(\pi/p)}{\pi/p}\right) > \alpha\frac{2c-3}{6}.$$

Then any set $A \subseteq \mathbb{Z}/p\mathbb{Z}$ of density at most α satisfying $|2A| < c|A| - 3$ is rectifiable.

Proof. First note that Theorem 5.3.1 tells us that $|\widehat{(1_A)}(\gamma)|$ achieves the value $\alpha\theta$, where

$$\theta = \sqrt{\frac{1-c\alpha}{c(1-\alpha)}}$$

for some $\gamma \in \widehat{\mathbb{Z}/p\mathbb{Z}}$. Then, by Corollary 5.2.10, we know that A has a large subset B of size M that is rectifiable, where we may take M any size up to

$$M_{max} := \frac{1}{2}|A| + \frac{p}{2\pi}\arcsin[\theta|A|\sin(\pi/p)].$$

Interestingly, to get the best bound in the proof, we won't choose M as large as possible. Instead, as any subset of B will still be rectifiable, we choose B so that $M > (c/3)|A|$ — this assumption is equivalent to the displayed equation in the statement of the theorem. Now consider Lemma 3.1.3 applied to B. It tells us that if $\mathrm{diam}(B) > 2M-2$ then

$$|2A| \geqslant |2B| \geqslant 3|B| - 3 > c|A| - 3$$

so we must have $\mathrm{diam}(B) \leqslant 2M-2$.

So now if we choose M such that $M > (c/3)|A|$ and $2M+2 < p/4$ then we may use the strong form of Lemma 5.4.4, which tells us that $\mathrm{diam}(A) < p/2$ so A is rectifiable also.

In particular, we can rectify A provided we can choose M in the range $(c/3)|A| < M \leqslant p/8$. We can find such an M provided $(c/3)|A| < p/8$, which is the condition we gave on α in the statement of the theorem.□

Corollary 5.4.6 should put this result into perspective.

Corollary 5.4.6. *Let $A \subseteq \mathbb{Z}/p\mathbb{Z}$ be a set of residues modulo a prime $p \geqslant 101$. If*

(i) *$c \leqslant 2.4$ and $\alpha < 3/34$; or*

(ii) *$c \leqslant 2.34$ and $\alpha < 3/8c$*

then if $|A| \leqslant \alpha p$ *and* $|2A| < c|A| - 3$ *then* $\operatorname{diam}(A) \leqslant |2A| - |A| + 1$.

Proof. One may check that the values of α and c specified satisfy Theorem 5.4.5. Then Theorem 3.1.1 finishes the result. □

Remark 5.4.7. Theorem 5.4.5 is not especially transparent. It is not exactly clear for which densities and which doublings the result holds, so Corollary 5.4.6 attempts to give context to the theorem in relation to the Freiman 2.4-Theorem. It should be noted, however, that the Freiman 2.4-Theorem itself was a simple corollary of a result known as the Freiman-Vosper theorem [N] which is very similar in spirit to Theorem 5.4.5. The statement of that result is as follows

Freiman-Vosper Theorem. Let $\alpha, c \in \mathbb{R}$ be such that $c \in [2, 3)$, $\alpha \in [0, 1/12)$ and

$$\frac{1 - c\alpha}{\sqrt{c}} < \frac{2c - 3}{3}.$$

Then any set $A \subseteq \mathbb{Z}/p\mathbb{Z}$ of density at most α satisfying $|2A| < c|A| - 3$ is rectifiable.

One may check that $\alpha = 1/35$ and $c = 2.4$ satisfy this theorem.

Aside from the obvious difference in the given bounds, the other main difference between Theorem 5.4.5 and the Freiman-Vosper theorem is the dependence on the prime p. The dependence on p is only slight, as the offending contributing factor

$$\frac{\sin(\pi/p)}{\pi/p}$$

tends quickly towards 1 as p gets large. Also, for example, the Freiman 2.4-Theorem only becomes non-trivial when $p > 105$ because otherwise the restriction on α means that A consists of two elements, and is hence an arithmetic progression for which the stated result is straightforward.

Thus putting $p \geqslant 101$ in Corollary 5.4.6 emphasises the improvement over the Freiman 2.4-Theorem as, even then, the result is non-trivial.

If one is interested in results that hold for p sufficiently large, one may shift the mentioned bounds in Corollary 5.4.6 a little bit further by taking limits with the bounds in Theorem 5.4.5, though there is limited interest in doing this, obviously.

Remark 5.4.8. The best known bounds in Freiman's 2.4-Theorem are due to Rodseth [R3], and our method of presentation does not appear to differ, morally, except that that our more algebraic exposition gives us the more general result of Theorem 5.4.5. This allows us more precise inferences than Rodseth's argument immediately allows. It should be noted, however, that Rodseth states that $\alpha < p/10.7$ will suffice provided $p \geqslant 139$, which lies slightly outside the range of α that our argument permits, thus there is some discrepancy between our results and thus there is likely a more subtle difference between our proofs that is not obvious to the author.

Despite this, curiously, the more general result of Theorem 5.4.5 actually represents an improvement over Rodseth's stated bound in the range $c < 2.35$, where $\alpha < 3/8c$ suffices. Rodseth's bounds have that $\alpha < 1/4c$ suffices for $c < 2.392$, so we have a definite and significant improvement over the range of densities for which the result holds. This is due to the careful selection of the size M of the set B in our proof that is does not come into play when considering c as large as possible.

One should note that, for instance, Theorem 5.4.5 becomes entirely ineffective when $c > 2.457$, approximately. At this point, the theorem offers no information whatsoever.

Remark 5.4.9. One can check, just as we did in Remark 5.4.2, that these results are still rectification results even though the interesting conclusion from these results is that sets of residues with small doubling have small diameter.

It is hence interesting to consider here an obvious case where this conclusion is true but where rectification is not possible; that is, where a set has small doubling and is contained within a relatively short arithmetic progression, but the set can almost certainly not be rectified at all.

Consider a set A of density larger than $1/c$. Then in any case $|2A| < c|A|$, simply because A is so big that it can't have any larger doubling. Furthermore, A itself is contained in an arithmetic progression of length at most $c|A|$ by the same reasoning, so A has relatively small diameter. Thus, a result akin to the Freiman 2.4-Theorem is trivially true for massive sets despite being almost certainly not rectifiable.

So consider the set $\Omega \subseteq [0,1] \times [2,3]$ consisting of all points (α, c) such that the Freiman 2.4-Theorem holds for a set A with density α and doubling at most c. This set can be vaguely split into two parts. The first part Ω_1 is where we can prove the set to be rectifiable on the basis of its small doubling. The second part Ω_2 is where the densities α are so large that the result holds for the trivial reasons mentioned above.

One can wonder what happens between the two sets. The obvious goal of the arguments we've seen is to prove that Ω_1 is as large as possible. But can Ω_1 be pushed so far that it closes the gap between itself and Ω_2? It seems unlikely. The arguments we have seen so far leave a wide gulf between Ω_1 and Ω_2. In fact, as mentioned in

the previous remark, Theorem 5.4.5 becomes entirely ineffective when $c > 2.457$, so for $c \in (2, 457, 3]$ the only rectification information results we can use are those of Freiman [F2] and Green-Ruzsa [GR], which have punishingly small densities.

So, a full characterisation akin to that of Vosper's Theorem or Theorem 4.3.1 of sets of residues with doubling between 2 and 3 is a long way off, so it appears there is large scope for improvement of rectification results based purely on small doubling, though likely new ideas are needed. To finish our discussion of rectification, we present a heuristic that the result ought to be attainable for sets A with density at most $1/4(c-1)$ for all c between 2 and 3.

5.5. **A Heuristic for Rectification.** In the proof of Theorem 5.4.5, one of the steps limiting the argument is that when we take the subset B we can only bound its doubling in terms of $|2A|$. It seems, to the author, to be a reasonable guess that one ought to be able to find a subset of a set with small doubling that itself has small doubling, and indeed if we are able to infer that the subset B has the same doubling c as that of A then we can make an improvement in the argument, as we now describe.

Start with a set A of density $\alpha > \beta$, where sets of density β and doubling c are rectifiable by, for example, Green-Ruzsa rectifiability [GR]. Then take a subset B of A of density βp and doubling c. By rectifiability,

$$\operatorname{diam}(B) \leqslant |2B| - |B| + 1 \leqslant |2A| - \beta p + 1 < c|A| - \beta p - 2 < (c\alpha - \beta)p$$

so provided that $c\alpha - \beta < 1/4$, B is rectifable. Furthermore, just as in the proof of Theorem 5.4.5, if $\beta > c\alpha/3$ then we can use the strong

form of the packing lemma to say that $\mathrm{diam}(A) < p/2$ and hence A is rectifiable (which proves the theorem). The two inequalities, paired together, just become

$$\alpha < \min\left\{\frac{3\beta}{c}, \frac{1}{4c} + \frac{\beta}{c}\right\}$$

Note that either gives an increase in the density over the density β of sets with doubling c that are rectifiable. Also, the leftmost bound is smaller than the rightmost bound unless $\beta > 1/8$. So now we iterate this argument:

Suppose that $\beta < 1/8$. Then iterating the argument n times means that sets of density $\beta(3/c)^n$ are rectifiable. Since $c < 3$, this eventually becomes more than $1/8$. Then the rightmost bound is smallest, so this is now what bounds the density increment. After m further iterations of this argument, we know that sets of density

$$\frac{1}{4c}\left(1 + \frac{1}{c} + \frac{1}{c^2} + \cdots + \frac{1}{c^m}\right) + \frac{\beta}{c^m}$$

as $k \to \infty$, this tends to $1/4(c-1)$.

Of course, this argument rests on the questionable assumption that we can find a subset of A with comparable doubling. If this can be proven, or the doubling of subsets can be bounded in some other way than using the doubling of the superset, then the above argument can be utilised as-is to make an improvement on Theorem 5.4.5.

6. Small Doubling in Binary Spaces

Now we have discuss sets with small doubling in the binary space $\mathbb{F}_2^n$, where throughout this section n is a positive integer. To get us started, we talk briefly about what we require from a Freiman theorem in a space with the lowest torsion possible.

Obviously, there is no point trying to show that a set A with small doubling is contained within a relatively short arithmetic progression because all arithmetic progressions in $\mathbb{F}_2^n$ have at most two elements, so this kind of statement would be fruitless and false. In view of the full Freiman theore, however, we can still talk about generalised arithmetic progressions here, though their structure falls into a somewhat more familiar type of set here in $\mathbb{F}_2^n$. Indeed, if it is not clear already, generalised arithmetic progressions in $\mathbb{F}_2^n$ are simply cosets of subspaces — the proof of this statement requires nothing more than to write out the definition of a generalised arithmetic progression in this setting.

So, we shall aim to show that sets of small doubling in $\mathbb{F}_2^n$ are contained within cosets of subgroups in $\mathbb{F}_2^n$.

We start our discussion with a result of Deshouillers, Hennecart and Plagne [DHP] and discuss its proof. We shall not present the entire proof of the theorem, but we shall discuss the proof and specifically how the bounds for the theorem come from it — the statement of the theorem involves a function that is not easily defined, but the explanation of where it comes from is more easily understood.

Theorem 6.0.1. *There is a function $u(x) : [1, 4] \to \mathbb{R}$ such that the following is true:*

Let A be a subset of the binary space $\mathbb{F}_2^n$. If $|2A| = c|A|$, where $c < 4$, then A is contained in the coset of a subgroup H of G such that

$$|H| \leqslant |A|/u(c).$$

A sizeable portion of the paper [DHP] is given entirely over to the discussion of the function u firstly to show that the function u does indeed satisfy the theorem and secondly in showing that for most of the range $c < 4$ this theorem represents the best known result for particularly small doubling in binary sets. We won't need to define the function u to discuss the above theorem, although one could indeed work out the function from our discussions.

We start out with a strong bound which is already very close to our target result. The proof comes straight from [DHP].

Theorem 6.0.2. *Let A be a subset of $\mathbb{F}_2^n$. If $|2A| = c|A|$ where $c < (3+\sqrt{5})/2$ then A is contained in a coset of a subgroup H of G such that*

$$|H| \leqslant \frac{2c-1}{-c^2+3c-1}|A|.$$

Proof. First, let *beta* $\in [0,2]$ and consider

$$\begin{aligned}\sum_{\gamma\neq\gamma_0} \widehat{1_A}(\gamma)\widehat{1_{2A}}(\gamma) &\leqslant \max_{\gamma\neq\gamma_0} |\widehat{1_A}(\gamma)|^{1-\beta}|\widehat{1_{2A}}(\gamma)|^{\beta} \left(\sum_{\gamma\neq\gamma_0} |\widehat{1_A}(\gamma)|^{1+\beta}|\widehat{1_{2A}}(\gamma)|^{1-\beta}\right)\\ &\leqslant \max_{\gamma\neq\gamma_0} |\widehat{1_A}(\gamma)|^{1-\beta}|\widehat{1_{2A}}(\gamma)|^{\beta}\\ &\quad\times \left(\sum_{\gamma\neq\gamma_0} |\widehat{1_A}(\gamma)|^2\right)^{(1-\beta)/2} \left(\sum_{\gamma\neq\gamma_0} |\widehat{1_{2A}}(\gamma)|^2\right)^{(1-\beta)/2}\end{aligned}$$

by Hölder's inequality, and hence

$$\sum_{\gamma\neq\gamma_0} \widehat{1_A}(\gamma)\widehat{1_{2A}}(\gamma) = \max_{\gamma\neq\gamma_0} |\widehat{1_A}(\gamma)|^{1-\beta}|\widehat{1_{2A}}(\gamma)|^{\beta} \times (\alpha(1-\alpha))^{(1-\beta)/2}(c\alpha(1-c\alpha)^{(1-\beta)/2}$$

Now we may use the weighted arithmetic-geometric mean inequality to note

$$\begin{aligned}|\widehat{1_A}(\gamma)|^{1-\beta}|\widehat{1_{2A}}(\gamma)|^{\beta} &\leqslant \beta^{\beta}(1-\beta)^{1-\beta}\left(|\widehat{1_A}(\gamma)| + |\widehat{1_{2A}}(\gamma)|\right) \\ &\leqslant \beta^{\beta}(1-\beta)^{1-\beta}\left(|\widehat{1_A}(\gamma)| + |\widehat{1_{2A}}(\gamma)|\right).\end{aligned}$$

so now we have the bound

$$\begin{aligned}\sum_{\gamma\neq\gamma_0}\widehat{1_A}(\gamma)\widehat{1_{2A}}(\gamma) \leqslant &\max_{\gamma\neq\gamma_0}\{|\widehat{1_A}(\gamma)| + |\widehat{1_{2A}}(\gamma)|\} \\ &\times \beta^{\beta}(1-\beta)^{1-\beta}(\alpha(1-\alpha))^{(1-\beta)/2}(c\alpha(1-c\alpha)^{(1-\beta)/2}.\end{aligned}$$

Considered as a single-variable function of β, a little calculus reveals this above equation achieves its maximum at

$$\beta = \frac{\sqrt{c\alpha(1-c\alpha)}}{\sqrt{\alpha(1-\alpha)} + \sqrt{c\alpha(1-c\alpha)}}$$

so that that

$$\beta^{\beta}(1-\beta)^{1-\beta}(\alpha(1-\alpha))^{(1-\beta)/2}(c\alpha(1-c\alpha)^{(1-\beta)/2} \leqslant \frac{\alpha(1-\alpha)\sqrt{c\alpha(1-c\alpha)}}{\sqrt{\alpha(1-\alpha)} + \sqrt{c\alpha(1-c\alpha)}}.$$

Now we consider

$$\max_{\gamma\neq\gamma_0}\{|\widehat{1_A}(\gamma)| + |\widehat{1_{2A}}(\gamma)|\}$$

Since γ is non-trivial, we may assume that γ is not constant on A. Moreover, as $\mathbb{F}_2^n$ has characteristic 2, all characters $\gamma \in \widehat{\mathbb{F}_2^n}$ satisfy

$\gamma(a) \in \{-1, 1\}$ for all $a \in A$. So we define the sets

$$A_1 = \{a \in A : \gamma(a) = 1\},$$

$$A_2 = \{a \in A : \gamma(a) = -1\},$$

$$B_1 = \{b \in 2A : \gamma(b) = 1\},$$

$$B_2 = \{b \in 2A : \gamma(b) = -1\}.$$

Notice that $B_1 = 2A_1 \cup 2A_2$ and $B_2 = A_1 + A_2$, so that each set above is non-empty and

$$\min\{|B_1|, |B_2|\} \geqslant \max\{|A_1|, |A_2|\}.$$

Furthermore, we have

$$\begin{aligned}
|\widehat{1_A}(\gamma)| + |\widehat{1_{2A}}(\gamma)| &= \frac{1}{|\mathbb{F}_2^n|}\left(|A_1| - |A_2|| + ||B_1| - |B_2||\right) \\
&= \frac{1}{|\mathbb{F}_2^n|}\left(2\max\{|A_1|, |A_2|\} - |A| + |2A| - 2\min\{|B_1|, |B_2|\}\right) \\
&\leqslant \frac{1}{|\mathbb{F}_2^n|}\left(|2A| - |A|\right),
\end{aligned}$$

where we have used the facts that $|A_1| + |A_2| = |A|$ and $|B_1| + |B_2| = |2A|$. Thus we have $\max_{\gamma \neq \gamma_0}\{|\widehat{1_A}(\gamma)| + |\widehat{1_{2A}}(\gamma)|\} \geqslant \alpha(c-1)$. Going back to our last bound on $\sum_{\gamma \neq \gamma_0} \widehat{1_A}(\gamma)\widehat{1_{2A}}(\gamma)$, we have hence shown that

$$\sum_{\gamma \neq \gamma_0} \widehat{1_A}(\gamma)\widehat{1_{2A}}(\gamma) \leqslant \alpha(c-1)\frac{\alpha(1-\alpha)\sqrt{c\alpha(1-c\alpha)}}{\sqrt{\alpha(1-\alpha)} + \sqrt{c\alpha(1-c\alpha)}}$$

However, if we note the following equality that was seen in Theorem 5.3.1

$$\alpha^2(1 - c\alpha) = \sum_{\gamma \neq \gamma_0} \widehat{1_A}(\gamma)\widehat{1_{2A}}(\gamma)$$

we have hence shown that

$$\alpha^2(1-c\alpha) \leqslant \alpha(c-1)\frac{\alpha(1-\alpha)\sqrt{ca(1-c\alpha)}}{\sqrt{\alpha(1-\alpha)}+\sqrt{c\alpha(1-c\alpha)}}.$$

After some algebraic manipulation, one may deduce that this indeed gives us the bound

$$\alpha = \frac{|A|}{|\mathbb{F}_2^n|} \geqslant \frac{-c^2+3c-1}{2c-1}.$$

In particular,

$$|\mathbb{F}_2^n| \geqslant \frac{2c-1}{-c^2+3c-1}|A|$$

Now a simple observation finishes off the proof: we may assume that A generates $\mathbb{F}_2^n$ so that the smallest subgroup of $\mathbb{F}_2^n$ containing A is $\mathbb{F}_2^n$ itself, and this last bound is exactly what we were after. □

Now we discuss how the improvement on this bound in the range $2.4 < c < 4$ is attained by using an 'induction on intervals' argument. We start with the interval $[1, c_1]$ where $c_1 = 2.4$ where the theorem holds for the function in Theorem 6.0.2. In the next step, we proceed to show that the theorem then holds for a weaker function when c is in an interval $[c_1, c_2]$, and then it holds for an even weaker function in the interval $[c_2, c_3]$, and so on. The limit of this sequence is c_k as $k \to \infty$ is 4. We then define u to be this function that is getting gradually weaker.

So now we explain the process that proves this inductive step works. As an inductive assumption, assume that c is in the interval $[u_k, u_{k+1}]$. Let γ denote a non-trivial character and, as in the proof above, let

$$A_1 = \{a \in A : \gamma(a) = 1\},$$

$$A_2 = \{a \in A : \gamma(a) = -1\}.$$

By translating if necessary, we may assume that $|A_1| > |A_2|$.

Now let H_1 denote the subgroup of $\mathbb{F}_2^n$ generated by A_1 and let S_2 denote the smallest subset of A_2 such that $A_2 \subseteq S_2 + H_0$. Then $|S_2|$ is the smallest number of cosets of H_1 that meets A_2, and hence

$$|A_1 + A_2| \geqslant |S_2||H_1| \geqslant |S_2||A_1|.$$

In particular, we have

$$|2A| = |2A_1| + |2A_2| + |A_1 + A_2| \geqslant |2A_1| + |S_2||A_1|.$$

If we let $d \in \mathbb{R}$ be the real number such that $|2A_1| = d|A_1|$

$$|2A| \geqslant (d + |S_2|)|A_1|$$

Now it may start to appear obvious how to set up the induction. We can find a character γ such that $|A_1|$ is very large, for example by using Theorem 5.3.1, which shows us in particular that we can't have both $|S_2|$ and d being large. It turns out that the essential case to consider is when both $|S_2|$ and d are simultaneously small, where it turns out small for d means $d \in [c_{k-|S_2|}, c_{k-|S_2|+1}]$ and hence we have $|H_1| \geqslant |A_1|/u(d)$.

Moreover, as $A = A_1 \cup A_2$ generates $\mathbb{F}_2^n$ we have $|\mathbb{F}_2^n| \leqslant 2^{|S_2|}|H_1|$, we also have

$$\alpha = \frac{|A|}{|\mathbb{F}_2^n|} \geqslant \frac{|A_1|}{2^{|S_2|}|H_1|} = \frac{u(d)}{2^{|S_2|}}.$$

To allow us to hazard a guess how u is defined, we assume that $|S_2| = 1$. Then we have $\alpha \geqslant u(d)/2$ where $d \in [c_{k-1}, c_k]$. In particular, we have shown that for $c \in [u_k, u_{k+1}]$ some $d \in [c_{k-1}, c_k]$ such that the theorem holds for $u(d)/2$, so one would guess to define $u(c) = u(d)/2$. Indeed, this is almost exactly how the function is defined in [DHP].

Remark 6.0.3. In their paper [DHP], Deshouillers, Hennecart and Plagne go through a few separate cases to show that if either $|S_2|$ or d are too large by themselves then the theorem is satisfied for more superficial reasons, and there is a fair amount of work in doing this. The point of this discussion is to show how the induction comes about and how the function u is defined, as this is the most interesting and elegant part of the proof, as far as the author is concerned.

6.1. **Small Doubling in Binary Spaces in the General Case.** Now we discuss briefly how these theorems look in the general case as c gets large.

There is a reasonably straightforward and well-known argument of Ruzsa, utilising only the celebrated Plünnecke-Ruzsa inequalities [TV], which proves the following.

Theorem 6.1.1 (Ruzsa)**.** *Let $A \subseteq \mathbb{F}_2^n$. If $|2A| < c|A|$ then A is contained within a subgroup H of G such that*

$$|H| \leqslant c2^{c^3-1}|A|.$$

A neat exposition of this argument is found in [TV], for example. The best result here in the general setting is due to Green and Tao [GT], where they show the following.

Theorem 6.1.2. *Let A be a subset of the binary space $\mathbb{F}_2^n$. If $|2A| < c|A|$ then there is a constant C such that A is contained in a subgroup H of G, where*

$$|H| \leqslant 2^{2c+C\sqrt{c}\log(c)}|A|.$$

The proof uses arguments from extremal set theory, which is apparently rare in additive combinatorics literature. It is in this setting

of $\mathbb{F}_2^n$ that we are closest to achieving the Polynomial Freiman-Ruzsa conjecture. See [GT] for a discussion.

7. Small Doubling in Non-Abelian Groups

Now we consider the analogue of small doubling in non-abelian groups. In our results so far, the use of commutativity has been constant and forgiving, yet when we move into an area without this valued abelian behaviour things get rapidly more complicated. While in the abelian case there is a well-known result of Kemperman [K] characterising subsets A of abelian groups G satisfying $|2A| < 2|A|$, there is no obvious corresponding result for non-abelian groups G.

We present here an English translation of a paper of of Freiman [F3], further modified by using the modern terminology of Additive Combinatorics. We try to follow Freiman's exposition as closely as possible, including his remarks, but the luxury of modern notation allows some of the results to be stated more clearly then Freiman was able and hence some of his exposition will be omitted as unnecessary.

We make a few remarks about notation. We shall continue to refer to the non-abelian analogue of doubling as such despite there being no addition taking place as this should help keep our discussion clear and consistent. We let the multiplicative dot $\cdot$ represent the group action of G, though as is customary we shall often leave the dot out as there shall be no ambiguity in doing so or we shall put it in to emphasisr certain multiplications. Also, we let $A^2 := A \cdot A$ denote the set of pairwise products of A in the obvious way. Further, we let A^{-1} denote the set of multiplicative inverses of elements of A and we shall let $H := H(A)$ denote the set $H := A^{-1} \cdot A$.

As said, the lack of commutativity is a big loss, as for instance in our definition of H we have to remember which side we multiply by reciprocals on. However, our first result in the study of non-abelian small

doubling shows that we are not completely hopeless: if the doubling of a set A is particularly small then A still retains at least the following abelian property.

Lemma 7.0.3. *Let A be a finite subset of a group G. If $|A^2| < 2|A|$ then*

$$A \cdot A^{-1} = A^{-1} \cdot A.$$

Proof. For all $a, b \in A$, notice that $|(a \cdot A) \cap (b \cdot A)| > 0$ so there is always an equality of the form $a^{-1}b = cd^{-1}$ for some $c, d \in A$. Hence $A^{-1} \cdot A \subseteq A \cdot A^{-1}$.

The reverse inclusion comes by considering the quantity $|(A \cdot a) \cap (A \cdot b)|$ for arbitrary $a, b \in A$ and utilising the same argument. □

So, though we've lost the abelian behaviour of elements, the product sets still retain some form of abelian behaviour when the doubling is small. If A has small doubling then $H = A^{-1} \cdot A$ and $H = A \cdot A^{-1}$ are equivalent definitions. In fact, as we proceed, we shall even find shortly that if the doubling is of particularly small order then we have some nice subgroup behaviour going on.

Lemma 7.0.4. *Let $A \subseteq G$ be a finite subset of the group G. If $|A^2| < \frac{1+\sqrt{5}}{2}|A|$ then there is an element $x \in A^2$ such that*

$$x \cdot H = H \cdot x = A^2.$$

Proof. First we set about finding an element $x \in A$ that will suffice. Let $\lambda := \frac{1+\sqrt{5}}{2}$ denote the golden ratio. Since

$$|A|^2 = \sum_{x \in A^2} |(A^{-1} \cdot x) \cap A|$$

there is an $x \in A^2$ such that $|(A^{-1} \cdot x) \cap A| > \frac{1}{\lambda}|A|$. Fix any such x.

Because the A has small doubling, notice that $|(A \cdot a) \cap (A \cdot b)| > (2-\lambda)|A|$ for all $a, b \in A$ so any element $y \in H$ satisfies

$$|(A \cdot y^{-1}) \cap A| > (2-\lambda)|A|.$$

This means that

$$|(A^{-1} \cdot x) \cap (A \cdot y^{-1})| > (2-\lambda+\frac{1}{\lambda})|A| - |A| = 0$$

so there is always an equality of the form $a^{-1} \cdot x = b \cdot y^{-1}$ for some $a, b \in A$. That is, $xy = ab \in A^2$ for every $y \in H$, so $x \cdot H \subseteq A^2$.

As an immediate consequence, notice that $|H| \leqslant |A^2| < \lambda|A|$; that is,

$$|A \cdot A^{-1}| < \lambda|A|$$

so the same argument as that above tells us that $|(A^{-1} \cdot a) \cap (A \cdot b^{-1})| > (2-\lambda)|A|$ for all $a, b \in A$, whence for any $z \in A^2$ we have $|(A^{-1} \cdot z) \cap A| > (2-\lambda)|A|$. Thus, with the same x as before, we have $|(A^{-1} \cdot x) \cap (A^{-1} \cdot z)| > 0$. This means that $yx^{-1} = ab^{-1} \in H$. In particular, $A^2 \cdot x \subseteq H$.

So we have that $|H| = |A^2|$ meaning that the two inclusions above are actually equalities; that is

$$A^2 = xH = Hx. \qquad \square$$

Remark 7.0.5. If it is not already clear, notice that the first step in the proof is a probabilistic argument. The average contribution from each summand is

$$\frac{|A|^2}{|A^2|} > \frac{|A|}{\lambda}$$

so there must be a summand achieving this average. We shall be using this style of probabilistic argument throughout the current section without further comment as it is will prove an incredibly useful tool.

Next, we move onto our first inverse theorem in a non-abelian setting.

Theorem 7.0.6. *Let $A \subseteq G$ be a finite subset of a group G. If $|A^2| < (3/2)|A|$ then either*

(i) *$A \subseteq H$, $A^2 = H$ and H is a subgroup of G; or*

(ii) *H is a normal subgroup of G, and there is an element $a \in A$ such that $A \subseteq a \cdot H$, $A^2 = a^2 \cdot H$*

Proof. Firstly, note that H contains its inverses. Secondly, for any $a, b \in A$ we have $|aA \cap bA| > \frac{1}{2}|A|$ and so $|xA \cap A| > \frac{1}{2}|A|$ for any $x \in H$. Thus, for any $x, y \in H$ we have $|x^{-1}A \cap yA| > 0$ and in particular there is always an equality of the form $xy = ab^{-1} \in H$ so H contains its products. This means that H is indeed a subgroup of G.

Now, let $a \in A$ be arbitrary. Then $A \subseteq aH$ and $A \subseteq Ha$, so $A \subseteq aH \cap Ha$. This means that

$$|(aHa^{-1}) \cap H| \geqslant |A| > \frac{2}{3}|H| > \frac{1}{2}|H|$$

and as $(aHa^{-1}) \cap H$ is a subgroup of H we have hence shown that $aHa^{-1} = H$. As we may assume without loss of generality that A generates G, then one of two things must be true: either $A \subseteq H$ or H is a normal subgroup of G. In the second case, as $A \subseteq aH$ for any $a \in A$ and $|A^2| = |H|$ by Lemma 7.0.4, we find $A^2 = a^2H$. □

7.1. **Slightly Larger Doubling.** As may be clear from the proof of Theorem 7.0.6, if the doubling of A is larger than $3/2$ then H may not be a subgroup of G at all. Instead, we may assert that H is the union

of a small number of cosets of some other subgroup J of G, where we define

$$J := \{g \in G : gH = Hg = H\}.$$

That H is a union of left-cosets of J is almost tautological, and the same can be said of right-cosets too. Indeed, we may write

$$H = \bigcup_{h \in H} hJ.$$

Hereon in, we shall retain this definition of J. Furthermore, we also retain the definitions of the integers m, n given by $n = |H|/|J|$ and $m = |AJ|/|J|$. We want to show that if the doubling of A is small enough then H is a union of a very small number of cosets of J. We require a single lemma.

Lemma 7.1.1. *Let $A \subseteq G$ be a finite subset of a group G. If $|A^2| \leqslant \mu|A| < \frac{1+\sqrt{5}}{2}|A|$ then*

$$|J| \geqslant \frac{1-\mu^2+\mu}{2-\mu}|A|.$$

Proof. We begin by making a couple of temporary definitions. We first let

$$r(h) := |Ah \cap A|$$

denote the number of representations of an element $h \in H$ as a ratio of elements $a^{-1}b$ of elements $a, b \in A$. Next, for $\alpha \in (0, 1)$, we let

$$R(\alpha) := \{h \in H : r(h) > \alpha|A|\}$$

denote the $\alpha|A|$-popular elements of H. We shall show that $R(\mu - 1)$ is both large and contained in J.

First, because $|Aa \cap Ab| \geqslant (2-\mu)|A|$ for all $a, b \in A$, we see that

$$r(h^{-1}) = |Ah^{-1} \cap A| \geqslant (2-\mu)|A|$$

for all $h \in H$. Thus, if $x \in H$ satisfies $rx = |Ax \cap A| > (\mu - 1)|A|$ then

$$|Ax \cap Ah^{-1}| > (2 - \mu + \mu - 1)|A| - |A| = 0$$

so there is an equality of the form $xh = a^{-1}b$. That is, $xH = H$. We may argue similarly to find that for such x we have $Hx = H$. Thus $R(\mu - 1) \subseteq J$.

Now we just need to show that $R(\mu - 1)$ is large. So let $\alpha \in (0, 1)$. Then

$$\begin{aligned} |A|^2 &= \sum_{h \in R(\alpha)} r(h) + \sum_{h \notin R(\alpha)} r(h) \\ &\leqslant |R(\alpha)| \times |A| + (|H| - |R(\alpha)|)\alpha|A| \end{aligned}$$

which we can rearrange to find

$$|R(\alpha)| \geqslant \frac{|A|^2 - \alpha|A||H|}{(1-\alpha)|H|} \geqslant \frac{1 - \alpha\mu}{1-\alpha}|A|,$$

whereupon the required bound follows by putting $\alpha = \mu - 1$. □

Corollary 7.1.2. *Let $A \subseteq G$ be a finite subset of a group G. If $|A^2| < \frac{8}{5}|A|$ then H is a union of at most 15 cosets of J.*

Proof. By Lemma 7.1.1, we know that $|J| > \frac{1}{10}|A|$, so

$$n = \frac{|H|}{|J|} < \frac{(8/5)|A|}{(1/10)|A|} = 16.$$ □

7.2. **Coset Culling.** At first sight, Corollary 7.1.2 tells us a lot of information about H when A has doubling at most 8/5. However, as

we shall go on to prove, the result overshoots somewhat. While it first tells us that H is a union of at most 15 cosets of J, it is in fact the case that H is a union of either 1 or 3 cosets of J. Thus we have to cull the other possibilities.

The claims we make henceforth shall apply equally to left-cosets and right-cosets, so we use the term coset in the general way to mean either and both — even though our proofs may be presented for left-cosets, for example, a near identical proof should convince the reader of the exact same fact for right-cosets.

A simple combinatorial lemma will provide us with all the ammunition we need to cull more than half of the possibilities from the list of supposedly attainable values for n.

Lemma 7.2.1. *Let $A, B, C \subseteq G$ be finite subsets of a group G. If $BA \subseteq C$ (resp. $AB \subseteq C$), $|A| > (2/3)|C|$ and $|B| > (1/2)|C|$ then $B^{-1}B$ (resp. BB^{-1}) is a subgroup of G.*

Proof. We present the proof only for the first claim, as the second claim follows an almost identical argument.

For all $b_1, b_2 \in B$ we have $|b_1A \cap b_2A| \geqslant 2|A| - |C|$ and consequently

$$|b_2^{-1}b_1A \cap b_3^{-1}b_4A| \geqslant 2(2|A| - |C|) - |A| > 0$$

for all elements $b_1, b_2, b_3, b_4 \in B$, so there is always an equality of the form $b_1^{-1}b_2b_3^{-1}b_4 = a_1a_2^{-1}$ for some elements $a_1, a_2 \in A$.

However, Ba_1 and Ba_2 always intersect. This tells us that there is always an equality of the form $a_1a_2^{-1} = b_5^{-1}b_6$ for some elements $b_5, b_6 \in B$. That is, $(B^{-1}B)^2 \subseteq B^{-1}B$, as required. □

With this lemma, we can attempt our first cull.

Lemma 7.2.2 (First Cull)**.** *Let $A \subseteq G$ be a finite subset of the group G. If*

$$|A^2| < \frac{8}{5}|A|$$

and $H \neq J$ then H is a union of either $3, 6, 9, 11, 12, 14$ or 15 cosets of J, in which case A is contained in a union of $2, 4, 6, 7, 8, 9$ or 10 cosets of J respectively.

Proof. First, note that

$$|AJ| = m|J| = \frac{m}{n}|H|$$

so in view of Lemma 7.2.1 it must be the case that $m/n \leqslant 2/3$, as otherwise we shall find that $A^{-1}A = H$ will be a subgroup of G, owing to the fact that $A \cdot AJ = xH$ for some $x \in A^2$.

Secondly, consider

$$|H| < \frac{8}{5}|A| \leqslant \frac{8}{5}|AJ| = \frac{8m}{5}|J| = \frac{8m}{5n}|H|.$$

This tells us that we need $8m/5n > 1$, and by reconciling with $m/n \leqslant 2/3$ we find that

$$m \in \left(\frac{5n}{8}, \frac{2n}{3}\right].$$

As m is an integer, we need this set to contain an integer. for the integers $n \leqslant 15$, it does so only when $n \in \{3, 6, 9, 11, 12, 14, 15\}$, and in each such case it contains only one integer: namely those mentioned in the statement of the theorem. □

Lemma 7.2.2 gives a small list of possible values of $|H|/|J|$ and $|AJ|/|J|$. To reduce this list of possibilities even further, we require another two technical results.

Lemma 7.2.3. *Let $A \subseteq G$ be a finite subset of the group G. If*

$$|A^2| < \frac{8}{5}|A|$$

and $H \neq J$ then $|AJ| < (16/15)|A|$ and $|JA| < (16/15)|A|$.

Proof. As $A^2 = xH$ is a union of cosets of J, we have $A \cdot AJ = A^2$. If it were the case that $|AJ| \geqslant (16/15)|A| > (2/3)|A^2|$ then Lemma 7.2.1 would tells us that $AA^{-1} = H$ is a subgroup of G, a contradiction. □

This simple observation allows us to leave only two remaining possibilities in the list.

Lemma 7.2.4. *Let $A \subseteq G$ be a finite subset of the group G. If*

$$|A^2| < \frac{8}{5}|A|$$

and $H \neq J$, then provided H is not a union of fifteen cosets of J, we have $AJ = JA$.

Proof. We suppose that $AJ \neq JA$. Then there is some element $a \in A$ such that $Ja \not\subseteq AJ$ and hence also such that

$$JaJ \not\subseteq AJ.$$

This means there is some $j \in J$ such that $jaJ \cap A = \emptyset$. In particular, we have

$$\begin{aligned} |JA| &= |JA \cap AJ| + |JA \cap (AJ)^C| \\ &\geqslant |A| + |JA \cap jaJ| \\ &\geqslant |A| + |ja \cap jaJ| \\ &= |A| + |A \cap aJ|. \end{aligned}$$

This means that

$$\begin{aligned} |A \cap aJ| = |A| - \left| \bigcup_{\substack{b \in A \\ b \notin aJ}} A \cap bJ \right| &\geqslant |A| - \left| \bigcup_{\substack{b \in A \\ b \notin aJ}} bJ \right| \\ &= |A| - (m-1)|J| \\ &= |A| - \frac{m-1}{n}|H| \\ &> \left(1 - \frac{8(m-1)}{5n}\right)|A| \end{aligned}$$

where $n = |H|/|J|$ and $m = |AJ|/|J|$. If we try the few possibilities for m and n given to us by Lemma 7.2.2 then apart from $n = 15$ this above amount exceeds $(1/15)|A|$ for all possible values of m and n, which is a contradiction in view of Lemma 7.2.3. □

Remark 7.2.5. If, in the proof of Lemma 7.2.4, we suppose that there are two left cosets $jaJ, j'aJ$ of J that do not intersect A then we can follow through the proof to get a contradiction even in the case when $n = 15$.

Also, notice that if JaJ, for some $a \in A$, contains a left-coset bH such that $bH \cap A = \emptyset$ but does not contain any right-cosets with the

same property, then the same argument as in Lemma 7.2.4 gives us a contradiction for all n also.

Finally, we note that the 'double cosets' JaJ for $a \in A$ which contain a left coset bH such that $bH \cap A = \emptyset$ is made up of two left cosets; if we suppose it is made up of three left cosets bH, cH, dH, say, then assuming without loss of generality that $|cH| > |dH|$ then the argument of Lemma 7.2.4 culminates in the fact that

$$|cH \cap A| \geqslant \frac{1}{2}\left(|A| - \frac{8(m-2)}{5n}|A|\right)$$

which provides the same contradiction as the original argument.

Now, we have two further culls of possibilities from the list, each of which wipe out all but one of the remaining possibilities. Thankfully, each leaves a different remaining possibility, so combining them shall give us the result we require.

Lemma 7.2.6 (Second Cull)**.** *Let $A \subseteq G$ be a finite subset of the group G. If*

$$|A^2| < \frac{8}{5}|A|$$

and $H \neq J$ then H is a union of either 3 or 15 cosets of J.

Proof. Lemma 7.2.4 tells us that

$$|(AJ)^2| = |JA \cdot AJ| = |A^2| < \frac{8}{5}|A| \leqslant \frac{8}{5}|AJ|$$

or, in particular

$$|(AJ)^2| < \frac{8}{5}|AJ|.$$

Moreover, as before with our original definition of H, we see that $(AJ)^{-1}AJ$ is a union of cosets of some subgroup K of G, where by

Lemma 7.1.1 we know that

$$|K| \geqslant \frac{1-(n/m)^2+(n/m)}{2-(n/m)}|AJ|.$$

and furthermore, by the same proof of Lemma 7.2.1, the number of cosets in this union is at most

$$t < \frac{n}{m} \times \frac{2-(n/m)}{1-(n/m)^2+(n/m)}$$

However, if we recall how we defined the subgroup J, then to make our subgroup K we need to consider the corresponding set $H' := (AJ)^{-1}(JA)$. But notice that

$$H' := (AJ)^{-1}(JA) = J^{-1}A^{-1}AJ = JHJ = H$$

by using Lemma 7.2.4, so it turns out $J = K$, and hence we must have $t = n$. However, for all admissable values of n, m apart from $n = 3, m = 2$ (not including $n = 15$ here, as Lemma 7.2.4 doesn't apply) it turns out that $t < n$, which is a contradiction. $\square$.

Lemma 7.2.7 (Third Cull)**.** *Let $A \subseteq G$ be a finite subset of the group G. If*

$$|A^2| < \frac{8}{5}|A|$$

and $H \neq J$ then H is a union of either 3 or 9 cosets of J.

Proof. Let $B = \{a_1, \ldots, a_m\}$ be elements of the m right cosets JA chosen such that they are contained in distinct left cosets of AJ — this is possible because any left-coset and any right-coset in the double-coset JaJ have non-empty intersection.

Now consider the products $a_i^{-1}a_j$ where $1 \leqslant i, j \leqslant m$ and $i \neq j$. There are $m(m-1)$ such products and they lie in $(n-1)$ left-cosets of

H, so some coset contains at least

$$q := \left\lceil \frac{m(m-1)}{n-1} \right\rceil$$

of these elements. Obviously, if $a_s^{-1}a_t$ and $a_u^{-1}a_v$ are in the same left-coset then $s \neq u$ otherwise we would have $a_t^{-1}a_v \in J$, which is false.

So there are q elements $h_1, \ldots, h_q \in H$ and numbers $1 \leqslant i_1, \ldots, i_q, j_1, \ldots, j_q \leqslant m$ such that

$$\begin{aligned} &a_{i_1}^{-1}a_{j_1}h_1 = a_{i_2}^{-1}a_{j_2}h_2 = \cdots = a_{i_1}^{-1}a_{j_1}h_1 \\ &= a_{i_1}^{-1}a'_{j_1} = a_{i_2}^{-1}a'_{j_2} = \cdots = a_{i_1}^{-1}a'_{j_1} \end{aligned}$$

where we have replaced the a_{j_k}s by elements a'_{j_k} chosen so that the left-cosets $a'_{j_k}J$ have non-empty intersection with A, which we made sure was possible by the second part of Remark 7.2.5.

Now $a_i^{-1}a_j = (ha_i)^{-1}(ha_j)$ for all $h \in J$, so the element $a_i^{-1}a_j$ has at least $q|H|$ representations of the form $b_i^{-1}b_j$ for some elements $b_i, b_j \in JA$. But we know that $|JA| < (m/n)|A^2|$, so

$$|JA \setminus A| < \frac{m}{n}|A^2| - |A|$$

and we can see that most of these $q|H|$ representations have $b_i, b_j \in A$. Specifically, the number of representations of $a_i^{-1}a_j$ of the form $b_i^{-1}b_j$ with $b_i, b_j \in A$ is at least

$$q|H| - 2\left(\frac{m}{n}|A^2| - |A|\right) > |A|\left(2 - \frac{|A|^2}{n|A|}(2m-q)\right).$$

However, one can check that for all remaining possibilities except $n = 9$, the above value exceeds $0.6|A|$, meaning that the specified elements are by definition elements of J, which is a contradiction. $\square$

Putting together the three culls, which amounted to a fair amount of effort, we have ruled out all but one possibility, as we desired.

Corollary 7.2.8. *Let $A \subseteq G$ be a finite subset of the group G. If*

$$|A^2| < \frac{8}{5}|A|$$

and $H \neq J$ then H is a union of 3 cosets of J.

7.3. **The Third Non-Abelian Inverse Theorem.** Corollary 7.2.8 is an inverse theorem as it stands, and we indeed consider it to be our second non-abelian inverse theorem, but it turns out we can do better than this. To state the theorem, we need first to go through a specific example of the types of set we have been considering. Freiman calls this result a theorem, so was likely at the time a new result due to him.

Example 7.3.1. Let $A \subseteq G$ be a set consisting of 4 elements such that $|A^2| = 6$, $H \neq J$ and J not a normal subgroup of G. We shall show that G is the semi-direct product of a group $\{1, h\}$ of order 2 with an abelian group that itself is the direct product of two cyclic groups with generators a, b satisfying the relations $hah = b$ and $hbh = a$. Furthermore, we show that $A = \{a, ah, ha, hah\}$.

Indeed, Lemma 7.2.1 tells us that $|J| \geqslant 2$, and as H is a union of 3 cosets of J by Corollary 7.2.8 we see that in fact $|J| = 2$. So assume that $J = \{1, h\}$ for some element $h \in G$ satisfying $h^2 = 1$.

We know that A is made up of two cosets of J, say

$$A = \{a, ah, c, ch\} = \{a, ha, c, hc\}.$$

We may assume that the left and right cosets don't coincide, because otherwise we would have $aJ = Ja$ and $cJ = Jc$ and J would be a

normal subgroup. Thus it turns out that $ah = hc$ and so

$$A = \{a, ah, ha, hah\}$$

and so

$$A^2 = \{a^2, a^2h, aha, ahah, ha^2h, haha, hahah\}$$

where we have duplicated some element. Ruling out elements that obviously can't be the same as one another, we are left with two possibilities, namely

(i) $aha = hahah$; or

(ii) $a^2 = ha^2h$.

If we let $c := ha$, say, and note that $h^2 = 1$ then the second possibility is easily seen to resemble the first. Finally, letting $b = hah$ gives us the structure we demanded at the start of this example.

With this example in hand, we are able to state our third, and final, non-abelian inverse theorem for sets with small doubling.

Theorem 7.3.2. *Let $A \subseteq G$ be a finite subset of at least three elements of the group G. If*

$$|A^2| < \frac{8}{5}|A|$$

then exactly one of the following cases holds:

(i) *H is a subgroup of G containing A, and $A^2 = H$;*

(ii) *H is a normal subgroup of G, and for any element $a \in A$, we have $A \subseteq aH$ and $A^2 = a^2H$;*

(iii) *J is a normal subgroup of G, A is contained in two cosets of J, and A^2 fills three costs of J;*

(iv) *there is a subgroup K of G such that A is contained in four cosets of K whose structure matches that of Example 7.3.1;*

(v) *there is a subgroup L of G such that $|L| = (1/3)|A^2|$, and there is some $a \in A$ such that $A \subseteq LaL$, $A^2 = La^2J$ and $|LaL| = (2/3)|A^2|$.*

Proof. If $J = H$ then we are in one of cases (i) or (ii), so we suppose that $J \neq H$. Then, by Corollary 7.2.8, H is a union of three cosets of J and A is contained in two cosets $aJ \cup bJ$, say, of J.

If either of a or b is in J then, by Lemma 7.2.4, J is a normal subgroup of G, and we are in case (iii), so we now rule this out.

In any case, it is still true that $aJ \cup bJ = Ja \cup Jb$. If $aJ = Ja$, then we are back in case (ii), so we again rule this out so there must be an element $a \in A$ such that $aJ \neq Ja$. Fix this element a.

Now as $AJ = JA$, there are elements $b, c \in G$ such that $AJ = aJ \cup bJ$ and $JA = Ja \cup Jc$. A fact we briefly mentioned in the proof of Lemma 7.2.7 is that we can choose the elements so that $b = c$. We go through the mentioned argument in this simpler case to show how it is done.

Indeed, if either $c \notin aJ$ or $b \notin Ja$ then there is obviously a choice that does the job. However, if $c = aj$ and $b = j'a$ for some $j, j' \in A$ then

$$j'ajJ = j'aJ = bJ$$

and

$$Jj'aj = Jaj = Jc$$

so we may replace b and c by jaj'. Thus we may assume that $A \subseteq aJ \cup bJ$ and $A \subseteq Ja \cup Jb$.

Now we define some more subgroups $J_1, J_2, J_3, J_4 \subseteq J$ of G by

(1) $aJ \cap Ja = aJ_1 = J_2a$; and

(2) $bJ \cap Jb = bJ_3 = J_4b$.

They are each subgroups of J because, for example, premultiplying $aJ \cap Ja = aJ_1$ by a^{-1} gives us $J_1 = J \cap (a^{-1}Ja)$.

Now note that

(3) $aJ \cap Jb = a(J \setminus J_1) = (J \setminus J_4)b$; and

(4) $bJ \cap Ja = b(J \setminus J_3) = (J \setminus J_2)a$.

If we consider

$$J_2(aJ \cap Jb) = (J_2a)J \cap (J_2J)b = (aJ_1)J \cap Jb = aJ \cap Jb$$

then (3) tells us that $(J \setminus J_4)b$ is closed under multiplication on the left by J_2. If $J \setminus J_4$ contained any element of J_2 it would follow that $1 \in J \setminus J_4$, which is false. Thus $J_2 \subseteq J_4$. Similarly, we find that $J_4 \subseteq J_2$, $J_1 \subseteq J_3$ and $J_3 \subseteq J_1$. That is, $J_1 = J_3$ and $J_2 = J_4$.

If $J_1 = J_2$ then J_1 is clearly a normal subgroup of G, and we are in case (iv), so we can assume that $J_1 \neq J_2$. In this case, by considering

(5) $aJ \cap Ja = aJ_1$; and

(6) $aJ \cap Jb = a(J \setminus J_1)$

and, in particular, multiplying on the left by J, we discover

$$JaJ \cap Ja = JaJ_1$$

and

$$JaJ \cap Jb = Ja(J \setminus J_1).$$

This means that

$$JaJ \cap (Ja \cup Jb) = JaJ$$

and we conclude $Ja \cup Jb \subseteq JaJ$. Also, by considering multiplying (5) and

$$bH \cap Ja = (J \setminus J_2)a$$

on the right by aJ, we see that

$$(bJaJ) \cap (Ja^2J) = (J \setminus J_2)a^2J$$

and

$$(aJaJ) \cap (Ja^2J) = J_2a^2J.$$

Putting these two bits together, we conclude that

$$Ja^2J \cap (bJaJ \cup aJaJ) = Ja^2J$$

or, completing the proof by showing we're in case (v), we have

$$A^2 = (bJaJ \cup aJaJ) = Ja^2J. \qquad \square$$

Remark 7.3.3. Some of the details in the proof above that we have meticulously explained were not explained so thoroughly in the paper of Freiman, so it is entirely possible that some of the argument can be shortened if there are simpler ways to deduce the same information.

That, however, concludes our discussion of small doubling in non-abelian groups. Arguably, the most inkeeping result to come from this section is Corollary 7.2.8. That is, the simplicity of that statement is more characteristic of the Freiman type results with which we are familiar, and the exacting breakdown of possible situations in Theorem 7.3.2 might be considered clutter in the realm of the neat qualitative results that we commonly seek.

8. Small Algebra Norm in Abelian Groups

In this section, we start our discussion of the algebra norm $\|1_A\|$ for subsets A of an abelian group G. Just as with sets with small doubling, one way to interpret this norm is as another measure of how grouplike a set A is in a manner we shall make precise soon enough. First, we'd better start by defining the algebra norm.

Definition 8.0.4. Let $f : G \to \mathbb{C}$ be a complex-valued function on the abelian group G. The *algebra norm* $\|f\|$ of f is defined to be

$$\|f\| = \sum_{\gamma \in \widehat{G}} |\widehat{f}(\gamma)|.$$

That is, the algebra norm is the L^1 norm of the Fourier transform of f.

We start by introducing some simple properties of the algebra norm. Firstly, we want to prove the multiplicative property that justifies the title 'algebra norm', so we need to study how the algebra norm behaves for functions multiplied together. So recall that for functions $f, g : G \to \mathbb{C}$ the function $f \cdot g : G \to \mathbb{C}$ is defined by $f \cdot g(x) = f(x)g(x)$ for each $x \in G$.

First, we need to prove a technical lemma regarding Fourier transforms of functions multiplied together. We have seen how the convolution of functions turns into multiplication of functions under the Fourier transform. Interestingly enough, multiplication of functions turns *almost* turns into convolution of functions under the Fourier transform too. If one considers the Inversion formula mentioned in §2.5 then one can see that Fourier inversion is itself almost exactly like a Fourier

transform. The reader can formalise this behaviour if he wishes, but in any case this insight makes it straightforward to prove the following.

Lemma 8.0.5. *If $f, g : G \to \mathbb{C}$ are complex-valued functions on the abelian group G then*

$$\widehat{f \cdot g}(\gamma) = \sum_{\eta \in \widehat{G}} \widehat{f}(\eta)\widehat{g}(\gamma - \eta)$$

for all $\gamma \in \widehat{G}$.

With that purely technical lemma out of the way, our first result justifying the term 'algebra norm' becomes obvious.

Lemma 8.0.6. *If $f, g : G \to \mathbb{C}$ are complex-valued functions on the abelian group G then*

$$\|f \cdot g\| \leqslant \|f\|\|g\|.$$

Proof. We have

$$\begin{aligned}\|f \cdot g\| = \sum_{\gamma \in \widehat{G}} \left|\widehat{f \cdot g}(\gamma)\right| &= \sum_{\gamma \in \widehat{G}} \left|\sum_{\eta \in \widehat{G}} \widehat{f}(\eta)\widehat{g}(\gamma - \eta)\right| \\ &\leqslant \sum_{\gamma, \eta \in \widehat{G}} |\widehat{f}(\eta)| \times |\widehat{g}(\gamma - \eta)| = \|f\|\|g\|. \qquad \square\end{aligned}$$

8.1. **The Algebra Norm as a Measure of Grouplikeness.** Now, we go on to explain how we can utilise the algebra norm as a measure of how grouplike a set is. What we mean is, for a subset $A \subseteq G$, the smaller $\|1_A\|$ is, the more like a a subgroup the set A is, in a manner we shall soon make precise. Note that if we talk of the algebra norm of a set A, we mean the algebra norm of its identity function 1_A. It shall be practical to do this frequently, so we make it a matter of definition.

Definition 8.1.1. Let $A \subseteq G$ be a subset of an abelian group G. Then we define the *algebra norm* $\|A\|$ of A to be

$$\|A\| := \|1_A\|$$

The first thing we note is the trivial consequence of Lemma 8.0.6, which offers a lower bound for sets under the algebra norm.

Corollary 8.1.2. *If $A \subset G$ is a nonempty subset of the abelian group G then $\|A\| \geqslant 1$.*

Proof. For any set A, $1_A \cdot 1_A = 1_A$, so Lemma 8.0.6 tells us that

$$\|A\| \leqslant \|A\|^2.$$

As A is non-empty, $\|A\| > 0$, so we are done. □

We claim that this lower bound is achieved if and only if the set A is the coset of a subgroup of G. To do this, we shall require a technical result concerning the Fourier transform.

Definition 8.1.3. Let $A \subseteq G$ be a non-empty subset of an abelian group G. We define the *orthogonal complement* $A^\perp$ of A to be

$$A^\perp = \{\gamma \in \widehat{G} : \gamma(x) = 1\}.$$

This is clearly a subgroup of $\widehat{G}$.

The reason we concern ourselves with the orthogonal complement is that it makes convenient the discussion of Fourier transforms of subgroups. In particular, the Fourier transform of a subgroup $H \leqslant G$ is the Haar measure of the orthogonal complement $H^\perp$. Indeed, from the

definition

$$\widehat{1_H}(\gamma) = \mathbb{E}_{x\in G}\, 1_H(x)\gamma(x) = \frac{1}{|G|}\sum_{x\in H}\gamma(x)$$

$$= \begin{cases} \dfrac{|H|}{|G|} & \gamma \in H^\perp \\ 0 & \gamma \notin H^\perp \end{cases}$$

(the orthogonality relations give the sum as 0 in the second case). Notice also that

$$|H^\perp| = |G|/|H|$$

by putting 1_H in Parseval's identity.

With this notation in mind, the following lemma is obvious. The corollary comes from pairing Lemma 8.0.6 with Lemma 8.1.4.

Lemma 8.1.4. *If $H \subset G$ is a subgroup of an abelian group G then for all $x \in G$ we have*

$$\|x + H\| = 1.$$

Corollary 8.1.5. *Let $A, H \subseteq G$ be non-empty subsets of an abelian group G, where $H \subseteq G$ is a subgroup of G, and let $x \in G$. If A intersects $x + H$ then*

$$\|A\| \geqslant \|A \cap (x + H)\|.$$

We want the converse of Lemma 8.1.4. The following curious result of Rudin [R1] allows us to do this.

Proposition 8.1.6. *Let $A \subseteq G$ be a nonempty subset of the abelian group G. If A is not a coset of G then*

$$\|A\| \geqslant \sqrt{5}/2.$$

Proof. As A is not a coset in G, there are elements $x, y, z \in A$ such that $x + y - z \notin A$. Define, for all $\gamma \in \widehat{G}$, the function

$$\begin{aligned} f(\gamma) :&= 2\gamma(z) + \gamma(y) + 2\gamma(x) - \gamma(x + y - z) \\ &= 2\gamma(z)(1 + \gamma(y - z)) + \gamma(x)(1 - \gamma(y - z)). \end{aligned}$$

The Fourier inversion formula tells us

$$1_A(a) = \sum_{\gamma \in \widehat{G}} \gamma(a)\widehat{1_A}(\gamma),$$

whence the first form of f shows us that $\sum_\gamma f(\gamma)\widehat{1_A}(\gamma) = 5$. However if we let $e^{2i\alpha} := \gamma(z - x)$, the second form of f reveals

$$|f(\gamma)| \leqslant 2|1 + e^{2i\alpha}| + |1 - e^{2i\alpha}| = 4|\cos(\alpha)| + 2|\sin(\alpha)| \leqslant 2\sqrt{5},$$

the last inequality coming from the fact that $4|\cos(\alpha)| + 2|\sin(\alpha)|$ will reach it's maximum on $[0, \pi/2]$, where the function is equal to $2\sqrt{5}\cos(\alpha + \theta)$ for some real θ. Hence we have, as if by magic,

$$5 = \sum_{\gamma \in \widehat{G}} f(\gamma)\widehat{1_A}(\gamma) \leqslant \|f\|_\infty \sum_{\gamma \in G} |\widehat{1_A}(\gamma)| \leqslant 2\sqrt{5}\|A\| \qquad \square$$

This gives us our first indication that the Algebra norm being small does indeed indicate that A is grouplike. This result of Rudin is interesting in its own right, and we discuss it later in §8.4. For now, we summarise what we have discovered in this section in the following Theorem.

Theorem 8.1.7. *Let $A \subseteq G$ be a non-empty subset of an abelian group G. Then A is a coset of a subgroup in G if and only if $\|A\| = 1$.*

The next section discusses the results of Cohen [C] and of Green-Sanders [GS2] which further demonstrate exactly how the algebra norm is a measure of grouplikeness.

8.2. **The Cohen and Green-Sanders Theorems.** As we promised, this section gives a result of Cohen which begins to justify our study of the algebra norm. We start with the resultant Green-Sanders theorem that is more relevant to our discussion, as we explain shortly.

Theorem 8.2.1 (Green-Sanders)**.** *Let $A \subseteq G$ be a subset of a finite abelian group G. If $\|A\| \leqslant M$ then we may write*

$$1_A := \sum_{j=1}^{L} \pm 1_{x+H_j}$$

where $L \leqslant \exp(\exp(CM^4))$ for some real number $C > 0$, $x_1, \ldots, x_L \in G$ and $H_1, \ldots, H_L$ are subgroups of G.

Furthermore, the number of distinct subgroups appearing in $H_1, \ldots, H_L$ is bounded above by $M + 1/100$.

This theorem is to the algebra norm as Freiman's theorem is to small doubling. More-or-less, it says that if the algebra norm of a set A is at most M then we can pick out M subgroups $H_1, \ldots, H_M$ and find that A a union of at most L cosets of these subgroups. Thus, the smaller M is, the more A appears like a subgroup of G, so it certainly justifies thinking of the algebra norm as a measure of grouplikeness.

The proof of Theorem 8.2.1 is long and complicated, and utilises a sophisticated 'Bohr Set' technology, so we do not go into it. A similar earlier result of Green-Sanders [GS1] for sets $A \subseteq \mathbb{F}_2^n$ mentions that the motivation for their method comes from Cohen's celebrated idempotent theorem, which we now state here in a rather non-standard form.

Theorem 8.2.2. *Let $f : G \to \mathbb{C}$ be a function on the abelian group G. Then $\|f\| < \infty$ if and only if f is the characteristic function of a set $A \subseteq G$ and takes the form*

$$f = \sum_{j \in J} \pm 1_{\gamma_j + \Gamma_j}$$

where J is finite.

The theorem is not normally stated in this way. Rather, it is a statement about idempotent measures. However, as an idempotent measure μ on $\widehat{G}$ is, as a matter of definition, a function such that $\widehat{\mu} \in \{0, 1\}$, stating it the above way is making the statement about the function $f = \widehat{\mu}$ rather than μ.

The problem with Cohen's Theorem that it is an entirely empty statement in finite abelian groups. The Green-Sanders Theorem attempts to bridge the gap so the statement has meaning in finite abelian groups by showing not only that $|J|$ is finite but that $|J|$ is bounded.

However, just as in Freiman's Theorem, they prove a massive bound on what is possible, namely the doubly-exponential bound stated in Theorem 8.2.1. If we want to prove statements about sets A that have $\|A\| < 2$, for example, where we might think we would get a precise answer, then the doubly-exponential bound might be disappointingly large.

Of course, for now, this is exactly the result we wanted. We have shown that the smaller the algebra norm $\|A\|$ is for a subset A of a finite abelian group, the more like a subgroup of G it is. So, in the next section, we consider a more specialist argument of Saeki [S2] that proves exactly the sort of result that is inkeeping with the theme

of our discussions, and characterises exactly sets A such that $\|A\| < (1+\sqrt{17})/4$.

8.3. **Saeki's Inverse Theorem.** This section will have us consider a result of Saeki that classifies a some subsets of groups in terms of the algebra norm of the set. To begin to prove the result and present it in the modern language of Additive Combinatorics will take some work, but to provide motivation we only need to go as for back as the Proposition 8.1.6 of Rudin.

Proposition 8.1.6 is a curious result. Viewed one way, it tells us that 1 is an isolated point in the image of sets $A \subseteq G$ under the algebra norm. That is to say, if the algebra norm $\|A\|$ of a set A satisfies $\|A\| < \sqrt{5}2$ then A is in fact a coset of a subgroup of G.

It turns out, as Saeki [S1] was able to prove, that we can prove this statement to holds up to $\|A\| < (1+\sqrt{2})/2$, and that result is best possible as, for example, the set $\{0,1\} \subseteq \mathbb{Z}/4\mathbb{Z}$ has $\|A\| = (1+\sqrt{2})/2$. Moreover, as the following lemma shows, once you go past this marker of $(1+\sqrt{2})/2$, there appear lots more sets with small algebra norm.

Lemma 8.3.1. *Let $A \subseteq G$ be a set of two elements of the abelian group G. Assume that $A = \{0, x\}$ where $x \in G$ is an element of order N. Then*

$$\|A\| = \begin{cases} \dfrac{2}{N\sin(\pi/2N)} & N \text{ odd} \\ \dfrac{2}{N\tan(\pi/2N)} & N \text{ even.} \end{cases}$$

Proof. Since $\widehat{1_A}(\gamma) = \frac{1}{|G|}(1+\gamma(x))$, summing over the cosets of $\langle x\rangle^\perp$ gives

$$\|A\| = \frac{1}{|G|}\frac{|G|}{N}\sum_{j=1}^{N}|1+e(j/N)| = \frac{2}{N}\sum_{j=1}^{N}|\cos(j\pi/N)|.$$

Using the symmetry about $\pi/2$, one evaluates this sum as the real part of the geometric series $\sum e(j/2N)$ to get the result. $\square$

For odd N, the value in the above Lemma decreases to $4/\pi$ as $N \to \infty$; for even N, it increases to the same value. The smallest value is obtained when $N = 4$, and is equal to $(1 + \sqrt{2})/2$. The largest value attained is $4/3$, which occurs when $N = 3$.

Notice however that, in the previous lemma, each set A is a union of two cosets of the trivial subgroup of G. It is straightforward to generalise this result as follows.

Lemma 8.3.2. *Let $A, H \subseteq G$ be subsets the abelian group G. Suppose that H is a subgroup of G and that $A = (x + H) \cup (y + H)$ for some $x, y \in G$. Let N denote the order of $x - y$ in G. Then*

$$\|A\| = \begin{cases} \frac{2}{N \sin(\pi/2N)} & N \text{ odd} \\ \frac{2}{N \tan(\pi/2N)} & N \text{ even.} \end{cases}$$

In particular, for any such A, $(1 + \sqrt{2})/2\|A\| \leqslant 4/3$.

Proof Without loss of generality, we take $y = 0$. Then, by definition,

$$\widehat{1_A}(\gamma) = \frac{1}{|G|} \left(\sum_{h \in H} \gamma(h) + \gamma(x + h) \right) = (1 + \gamma(x)) \widehat{1_H}(\gamma)$$

However, remembering that $\widehat{1_H}$ is the Haar measure of $H^\perp$, we discover

$$\|A\| = \mathbb{E}_{\gamma \in H^\perp} |1 + \gamma(x)|.$$

This really is the same sum as in Lemma 8.3.1, where instead of summing over the group $G^\perp$ we're now summing over $H^\perp$, which is canonically isomorphic to $\widehat{G/H}$ $\square$

If the reader is unfamiliar with the orthogonal complement, he can check that last statement of this proof is obvious by constructing the obvious isomorphism. Otherwise, one can check out the book of Tao and Vu for further discussion on this [TV].

In view of Lemma 8.3.2, now seems a good time to state the theorem of Saeki that we are aiming for

Theorem 8.3.3 (Saeki)**.** *Let $A \subseteq G$ be a subset of an abelian group G. If $1 < \|A\| \leqslant (1+\sqrt{17})/4$ then A is a union of two cosets of some subgroup of G.*

This theorem says that the examples given in Lemma 8.3.2 of sets with

$$\|A\| \leqslant (1+\sqrt{17})/4$$

are the only examples. Notice that $(1+\sqrt{17})/4 < 4/3$, so there is room for improvement on this result, and indeed Saeki goes so far as to conjecture that the result ought to hold for all A satisfying $\|A\| < 4/3$. Indeed, in the direction of proving this conjecture, we shall show that the result holds for up to the value $\|A\| < 9/7$ by carefully refining one of Saeki's lemmas.

If we take our earlier analogy a little further in saying that Theorem 8.2.1 is to the algebra norm as Freiman's theorem is to small doubling, then Cohen's theorem is the Cauchy-Davenport theorem of the algebra norm, and Saeki's result is the Freiman $3k-3$ Theorem of the algebra norm.

Saeki's Theorem is a careful improvement to Rudin's result Proposition 8.1.6. The magic of that result was the selection of a particular

function f that we could pit against the set A to prove the lower bound. The next section tries to put this function f into context.

8.4. **Saeki Functions.** The function f from the proof of Proposition 8.1.6 looks in some sense like the Fourier transform of a set. These functions allow us to change combinatorial information into analytic information, as we now consider.

Throughout this section, we consider A to be fixed in G. Then, for any non-zero complex-valued function f on G, it is obvious that

$$\|A\| \geqslant \frac{1}{\max_\gamma |f(\gamma)|} \left| \sum_\gamma f(\gamma)\widehat{1_A}(\gamma) \right|$$

and if this rightmost ratio is bigger than 1 then we get a nontrivial lower bound on $\|A\|$. We now give a name to such functions.

Definition 8.4.1. Let $A \subseteq G$ be a subset of an abelian group G, and let $f : G \to \mathbb{C}$ be a complex-valued function on G. If the ratio

$$S(f) := \frac{1}{\max_\gamma |f(\gamma)|} \left| \sum_\gamma f(\gamma)\widehat{1_A}(\gamma) \right|$$

satisfies $S(f) > 1$, then we say that f is *Saeki* and call $S(f)$ the *Saeki constant* of f.

We have seen one Saeki function so far, and we quickly examine it to show that, in some sense, it is best possible.

Example 8.4.2. Rudin's function has the form

$$f(\gamma) = \lambda\gamma(x)[1 + \gamma(z - x)] + \gamma(y)[1 - \gamma(x - z)]$$

where $x, y, z \in A$ are chosen to satisfy $x+y-z \notin A$, yielding $\sum_\gamma f(\gamma)\widehat{1_A}(\gamma) = 2\lambda + 1$. Also, as in the proof from Rudin's book, there is some α such

that

$$|f(\gamma)| \leqslant 2\lambda|\cos(\alpha)| + 2|\sin(\alpha)| \leqslant 2\sqrt{\lambda^2+1}|\cos(\alpha+\theta)|$$

and we hence find that

$$S(f) \geqslant \frac{1+2\lambda}{2\sqrt{1+\lambda^2}}$$

which, elementary calculus shows, takes its maximum at $\lambda = 2$, as was chosen by Rudin. Thus one can not hope to do any better with a function f given by the above form.

Saeki's improvement to this result essentially lies in finding more suitable functions that make use of more combinatorial information. Saeki uses a different Saeki function in each of his two papers [S1, S2], namely one of the two forms

$$\lambda\gamma(z)[1+\Re\gamma(y-z)] + \gamma(x)[1-\Re\gamma(y-z)],$$

$$\lambda\gamma(z)[1+\gamma(y-z)] + \gamma(x)[1-\Re\gamma(y-z)]$$

for some $x, y, z \in A$ satisfying some further additive conditions.

The Saeki function that gave Saeki the best results in [S2] is the second of those above.

Lemma 8.4.3. *Let $A \subseteq G$ be a subset of an abelian group G. If there exist $x, y, z \in A$ such that $x+y-z, x-y+z \notin A$. Then*

$$\|A\| \geqslant (1+\sqrt{17})/4.$$

Proof. For $\lambda > 0$, put

$$\begin{aligned} f(\gamma) =& \lambda\gamma(z)[1+\gamma(y-z)] + \gamma(x)[1-\Re\gamma(y-z)] \\ =& \lambda[\gamma(z)+\gamma(y-z)] \\ &+ \gamma(x) - \tfrac{1}{2}[\gamma(x+y-z)+\gamma(x-y+z)] \end{aligned}$$

The second expression shows that $\sum_\gamma f(\gamma)\widehat{1_A}(\gamma) = 2\lambda + 1$, and, upon choice of α so that $\gamma(y-z) = e^{2i\alpha}$, the first expression shows

$$\begin{aligned} |f(\gamma)| &\leqslant \lambda|1+e^{2i\alpha}| + |1-\cos(2\alpha)| \\ &= 2\lambda|\cos(\alpha)| + 2(1-\cos^2(\alpha)) \\ &= \tfrac{1}{2}(\lambda^2+4) - 2(|\cos(\alpha)| - \lambda/2)^2 \leqslant \tfrac{1}{2}(\lambda^2+4) \end{aligned}$$

so that $\|A\| \geqslant (4\lambda+2)/(\lambda^2+4)$, the maximum of which is attained at $\lambda = (\sqrt{17}-1)/2$ where it takes the value $(1+\sqrt{17})/4$. □

The next lemma characterises when the other Saeki function mentioned above can be utilised. One of the instances of its good use is superceeded by the above lemma, but one of its instances proves useful in the binary space $\mathbb{F}_2^n$, where n is a positive integer.

Lemma 8.4.4. *Let $A \subseteq G$ be a subset of an abelian group G. If there exist $x, y, z \in A$ such that $x+y-z, x-y+z \notin A$ then $\|A\| \geqslant 5/4$. If further $2z-y \in A$, then $\|A\| \geqslant 3/2$.*

Proof For $\lambda \geqslant 1$ (if we put $\lambda < 1$ and go through the following, we get a trivial result), let

$$\begin{aligned} f(\gamma) =& \lambda\gamma(z)[1 + \Re\gamma(y-z)] + \gamma(x)[1 - \Re\gamma(y-z)] \\ =& \lambda[\gamma(z) + \tfrac{1}{2}\gamma(y-z) + \gamma(2z-y)] \\ &+ \gamma(x) - \tfrac{1}{2}[\gamma(x+y-z) + \gamma(x-y+z)] \end{aligned}$$

Letting $\delta = 1_A(2z-y)$, the second expressions shows us $\sum_\gamma f(\gamma)\widehat{1_A}(\gamma) = \frac{3+\delta}{2}\lambda + 1$, and because $\lambda \geqslant 1$, the first expression tells us $|f(\gamma)| = \lambda|1 + \Re(w)| + |1 - \Re(w)| \leqslant 2\lambda$, where $w = \gamma(y-z)$. We thus discover

$$\|A\| \geqslant \frac{(3+\delta)\lambda + 2}{4\lambda}.$$

Bearing in mind that $\lambda \geqslant 1$, the above takes its max at $\lambda = 1$. If $\delta = 1$, this maximum is $3/2$; if not, the maximum is $5/4$. □

It is useful, however, in $\mathbb{F}_2^n$; since addition there is the same as subtraction (hence negatives the same as positives) and $2y = 0$ for all $y \in \mathbb{F}_2^n$. In particular, we immediately gain the following result.

Corollary 8.4.5. *If $A \subseteq \mathbb{F}_2^n$ is not a coset, then $\|A\| \geqslant 3/2$.*

Proof. Since A is not a coset, there are $x, y, z \in A$ such that $x + y - z = x - y + z \notin A$ and $2y - z = z \in A$. Applying the previous lemma gives the result. □

This result is proved by a near-identical method in [GS1].

It turns out that, if we are careful, we can do better than Lemma 8.4.3. This argument is due to the author.

Lemma 8.4.6. *Let A be a subset of the finite abelian group G. Suppose there are elements $x, y, z \in A$ such that the two elements $x+y-z$ and $x-y+z$ are not contained in A. Then $\|\widehat{1_A}\|_{L^1} \geqslant 9/7$.*

Proof. We let $\Re(z)$ denote the real part of the complex number z, as is standard. Now we define a function $f : \widehat{G} \to \mathbb{C}$ by

$$\begin{aligned} f(\gamma) &= \mu(\theta - \Re\gamma(y-z))\gamma(x) + \gamma(z)(1+\gamma(y-z)) \\ &= \mu\theta\gamma(x) + \gamma(y) + \gamma(z) - \frac{\mu}{2}(\gamma(x+y-z) + \gamma(x-y+z)) \end{aligned}$$

for each $\gamma \in \Gamma$, where we take μ, θ to be constants with $0 < \mu, \theta < 1$. The second form in which we've written f shows us that

$$\int_{\gamma \in \widehat{G}} \widehat{1_A}(\gamma) f(\gamma) = \mu\theta + 2$$

For any given γ, we know that $\gamma(y-z) = e^{2i\alpha}$ for some $\alpha \in \mathbb{R}$. Thus the first form we've written f illustrates that, for any $\gamma \in \widehat{G}$,

$$\begin{aligned} |f(\gamma)| &\leqslant \mu|\theta - \cos(2\alpha)| + |1 + e^{2i\alpha}| \\ &= \mu|\theta - 2\cos^2(\alpha) + 1| + |e^{-i\alpha} + e^{i\alpha}| \\ &= 2\mu\left|\frac{1+\theta}{2} - |\cos(\alpha)|^2\right| + 2|\cos(\alpha)| \end{aligned}$$

Now we consider two cases. The first case is if $|\cos(\alpha)|^2 < \frac{1+\theta}{2}$. In this case, we notice

$$\begin{aligned} |f(\gamma)| &\leqslant 2\mu\frac{1+\theta}{2} - 2\mu|\cos(\alpha)|^2 + 2|\cos(\alpha)| \\ &= -2\mu\left(|\cos(\alpha)|^2 - \frac{1}{2\mu}\right)^2 + \frac{1}{2\mu} + \mu(1+\theta) \\ &\leqslant \frac{1}{2\mu} + \mu(1+\theta). \end{aligned}$$

On the other hand, if we have $\frac{1+\theta}{2}| \leqslant \cos(\alpha)|^2 \leqslant 1$, we discover

$$|f(\gamma)| \leqslant 2\mu\left(1 - \frac{1+\theta}{2}\right) + 2 = 2 + \mu(1-\theta).$$

Summarising, we thus have

$$\|f\|_\infty \leqslant \max\left\{\frac{1}{2\mu} + \mu(1+\theta), 2 + \mu(1-\theta)\right\}$$

To maximise this estimate, we need only minimise our estimate for $\|f\|_\infty$, so we now consider this. Indeed, we may write it as

$$\|f\|_\infty \leqslant \left\{\frac{1}{2\mu} + \mu(1+\theta), 2 + \mu(1-\theta)\right\} = 2+\mu(1-\theta)+\max\left\{\frac{1}{2\mu} + 2\mu\theta - 2, 0\right\}$$

and this rightmost term is zero provided that

$$\theta \leqslant \frac{1+4\mu}{4\mu^2}.$$

Since our estimate for $\|f\|_\infty$ clear increases as θ decreases, we take the maximum possible value of θ, namely $\theta = \frac{1+4\mu}{4\mu^2}$. It is then a straightforward application of elementary calculus optimising for μ, and we find that $\mu = 2/3$ optimises our bound, yielding $\theta = 15/16$. Hence, as before, we deduce

$$\|\widehat{1_A}\|_{L^1} \geqslant \frac{|\int_{\gamma\in\widehat{G}} \widehat{1_A}(\gamma)f(\gamma)|}{\|f\|_\infty} \geqslant (2+\mu\theta) \Big/ \max\left\{\frac{1}{2\mu} + \mu(1+\theta), 2 + \mu(1-\theta)\right\}$$

which tells us

$$\|\widehat{1_A}\|_{L^1} \geqslant \left(2 + \frac{2}{3}\times\frac{15}{16}\right) \Big/ \left(2 + \frac{2}{3}\left(1 - \frac{15}{16}\right)\right) = \frac{9}{7}. \qquad \square$$

8.5. **Combinatorial Information for Saeki Functions.** Saeki's Inverse Theorem is a strong structural statement. It states that if $\|A\|$ is small enough, then A is a union of at most two cosets in G, and the proof utilises Lemma 8.4.3, which tells us that if we can find a certain triple of points in our set A, then A has a large algebra norm. Thus

our discussion turns to categorising sets that fit this combinatorial information.

For Rudin, this deduction was easy, as one can distinguish cosets in the following way: every triple of points $\{x, y, z\} \subseteq A$ satisfies $x + y - z, x - y + z, -x + y + z \in A$ if and only if A is a coset in G. Thus if A is not a coset, then you can find a triple of points that will work in Rudin's function. Saeki's function amplifies how errant a triple of points of A can be, and Lemma 8.4.3 hints that its counterpart will be a structure theorem for sets that contain at least 2 of the sums $x + y - z, x - y + z, -x + y + z \in A$ for any given $x, y, z \in A$. Given this requirement, we call a triple of points $\{x, y, z\} \subseteq A$ *errant* if

$$\{x + y - z, x - y + z, -x + y + z\} \cap A \leqslant 1.$$

A careful examination of Saeki's inverse theorem shows that the proof is a structure theorem for sets that have no errant triples. To state this reformulation, we introduce the notion of a punctured coset. A *punctured coset of* H in G is a coset of a subgroup of $H \leqslant G$ with a point omitted $h \in H$.

Lemma 8.5.1. *Suppose that A contains no errant triples. Then either A is a union of two cosets of some coset in G, or A contains a punctured coset of H, where H is a cyclic subgroup of G with odd order.*

The proof is a case by case analysis built on an induction. Hence we start with the following initial case.

Lemma 8.5.2. *If $A = \{x, y, z\}$ is not a coset, then A is an errant triple.*

Proof. If A contains neither $x+y-z$ nor $x-y+z$, then we have an errant triple. Hence it must contain at least one of these elements. If both $x+y-z, x-y+z$ are in A, then we must have $x+y-z=z$ and $x-y+z=y$ whence $y-z=z-x$ and $x-y=y-z$. This would mean that $x-y=z-x$ and so $-x+y+z=x\in A$, and hence that A is a coset, which is false. Thus, by symmetry, A can contain at most one element out of each of the pairs $\{x+y-z, x, x-y+z\}, \{x+y-z, -x+y+z\}, \{x-y+z, -x+y+z\}$. Consequently $\{x, y, z\}$ is an errant triple. □

We move onto a theorem very close to our target result.

Theorem 8.5.3. *If $|A|\geqslant 3$ and A is not a union of proper nontrivial cosets of some subgroup of G, then either A contains an errant triple, or A contains a punctured coset of H, where H is cyclic of odd order.*

Proof. The proof is an induction on $|A|$, where the case $|A|=3$ comes immediately from Lemma 8.5.2 using the assumption that A is not a coset. In the following, we assume that $|A|\geqslant 4$ and assume inductively that the result hold for all sets with lesser size. Note that we only need our inductive assumption in the first of the following cases.

Case 8.5.3.1. A contains a coset of $\langle y\rangle$ where $(y\in G)$.

Without loss of generality, we assume that y has prime order and $\langle y\rangle\subset A$. As A is not a union of cosets, there is some $x\in A$ such that $x+\langle y\rangle\not\subseteq A$. Then $A'=A\cap(x+\langle y\rangle)$ is strictly smaller than A, and we can use the inductive assumption provided $|A'|\geqslant 3$, hence it remains to consider when $|A'|\leqslant 2$.

Note that $x \in A'$. If neither $x+y$ nor $x-y$ are contained in A', we have the errant triple $\{x, 0, y\}$ (notice that this covers the two cases $|A'| = 1$ and when the order of y is 2). Not both $x+y, x-y$ can be in A' since neither is equal to x. Thus exactly one of $x+y, x-y$ is in A' and because $\langle y \rangle = \langle -y \rangle$, we may assume that it is $x+y$.

Since A' is not a coset, y has order greater than 2. This means that $x - y \notin A'$ (if it was, it would mean $x + y = x - y$ and thus $2y = 0$) and $x + 2y \notin A'$ by identical reasoning. If we know further that the order of y is bigger than 3, then $x - 2y \notin A'$ by the same reasoning and we have the errant triple $\{x, 0, 2y\}$, so we are left to consider the case when y has order 3. But, in this case, $A' = \{x, x+y\} = x + \{0, y\}$ is a punctured coset.

Case 8.5.3.2. A contains a 3-AP, but not a coset.

Without loss of generality, we assume the 3-term progression contained in A is $\{0, y, 2y\}$ and, as A does not contain $\langle y \rangle$, y has order at least 4.

Partition $A \cap \langle y \rangle$ into a family of sets $\mathcal{B}$ of the form $\{ny, (n+1)y, \ldots, (n+N)y\}$ (some n, N) where $(n-1)y, (n+N+1)y \notin A$. One of the sets $B \in \mathcal{B}$ has size at least three (in particular, the one containing $0, y, 2y$ does, and possibly more do too).

If $\mathcal{B}$ contains an element $B = \{ny, (n+1)y, \ldots, (n+2k)y\}$ with odd size then we have the errant triple $\{ny, (n+k)y, (n+k+1)y\}$ so we henceforth assume that all members of $\mathcal{B}$ have even size. If $\mathcal{B}$ contains an element $B = \{ny, (n+1)y, \ldots, (n+2k-1)y\}$ (where now $2k-1 \geqslant 3$) with $(n+2k+1)y \notin A$, then we have the same errant triple $\{ny, (n+k)y, (n+k+1)y\}$.

We call the elements in $\langle y \rangle$ that aren't in $A \cap \langle y \rangle$ the 'gaps'. The previous statement implies that all gaps have length 1. If there exist more than one gap, then there are two elements $B_1, B_2 \in \mathcal{B}$ such that

$$B_1 = \{ny, \ldots, (n+2k-1)y\}$$

$$B_2 = \{(n+2k+1)y, \ldots, (n+2k+2j)y\}$$

with $|B_1| \geqslant |B_2|$; that is, $k \geqslant j$. In this case, either $\{(n+k+j)y, ny, (n+k+j+1)y\}$ or $\{(n+k+j)y, (n+1)y, (n+k+j+2)y\}$ make up an errant triple. Thus we may assume that there is only one gap and that it has length 1. But this says exactly that there is a punctured coset of H in A, and notice that H is cyclic with odd order.

This leaves us with a final case to consider, which we shall quickly dispatch.

Case 8.5.3.3. A contains no 3-APs

In this case, any three distinct elements $x, y, z \in A$ make up an errant triple. If, for example, $x + y - z \in A$ then $x - y + z \notin A$ else it would complete the 3-AP $(x + y - z, x, x - y + z)$. Thus, by symmetry, A can contain at most one element out of each of the pairs $\{x+y-z, x, x-y+z\}, \{x+y-z, -x+y+z\}, \{x-y+z, -x+y+z\}$. Consequently $\{x, y, z\}$ is an errant triple. □

8.6. **The Proof of Saeki's Inverse Theorem.** Now we are able to put the pieces together to get a result of which the earlier mentioned result of Saeki is a straightforward Corollary.

Corollary 8.6.1. *Let $A \subseteq G$ be a subset of an abelian group G. If $|A| \geqslant 3$ and A is not a union of cosets of some non-trivial subgroup of G then $\|A\| \geqslant (1 + \sqrt{17})/4$.*

Proof. This comes immediately from the previous theorem. If A has an errant triple, then we use Lemma 8.4.3. If instead it contains A', a punctured coset of a cyclic subgroup with odd order N, then we apply Corollary 8.1.5 as follows.

We may now consider A' to be a subset of the cyclic group $\mathbb{Z}/N\mathbb{Z}$, and as such all the size of all the non-trivial Fourier coefficients of A' are the same as that of a singleton, namely $1/N$. The trivial Fourier coefficient contributes the density of A', namely $(N-1)/N$, hence A' has algebra norm $2(N-1)/N = 2 - 2/N$. As $N \geqslant 3$, this is at least $4/3 > (1+\sqrt{17})/4$. □

Note that Lemma 8.4.3 was used in the above proof to complete our discussion of Saeki's proof of Theorem 8.3.3. The reader should notice we could just as easily have used Lemma 8.4.6, so the following needs no proof.

Corollary 8.6.2. *Let $A \subseteq G$ be a subset of an abelian group G. If $|A| \geqslant 3$ and A is not a union of cosets of some non-trivial subgroup of G then $\|A\| \geqslant 9/7$*

Finally, this gives us our improvement to Saeki's Theorem. If the fact that Saeki's Theorem is a immediate consequence of Corollary 8.6.1 is not obvious to the reader, the following proof spells it out.

Corollary 8.6.3. *If $1 < \|A\| < 9/7$ then A is a union of two cosets of some subgroup of G.*

Proof. We prove this by induction on $|A|$. When $|A| = 2$, A is always a union of two cosets of the trivial subgroup, so we're done, and we henceforth suppose that $|A| \geqslant 3$.

Now we apply Corollary 8.6.2. Because A fails the size bound for the algebra norm, we must have that A is a union of cosets of some subgroup H. But then the algebra norm of A is the same as the set $A+H$ considered as a subgroup of G/H. By the inductive hypothesis, $A+H$ is a union of two cosets, and hence so too is A. $\square$

Remark 8.6.4. Saeki's papers look very different to the arguments we have employed here, but they are essentially the same. Just as with Cohen's Theorem earlier, where we changed the vantage point of the theorem from the group G to the dual group $\widehat{G}$ to get an equivalent result, we have done the same thing with Saeki's Theorem also.

The reasons for doing this should make themselves obvious. Firstly and foremostly, this presentation is inkeeping with our arguments and helps maintain the flow of discussion. Secondly, presenting the arguments in this manner allow the use of tools of Additive Combinatorics which would otherwise have been more cumbersome to utilise.

Remark 8.6.5. Saeki conjectured that Corollary 8.6.1 should hold for sets $A \subseteq G$ satisfying $\|A\| < 4/3$, and we see no reason why this conjecture would be obviously false. Furthermore, in researching the attained improvement to Saeki's Theorem, the author did various numerical tests for varying Saeki functions that are not here discussed. That testing indeed suggested that $4/3$ is the correct target, but the problem is rather not in coming up with these Saeki functions but rather in proving the complementary structural theorems that make the Saeki functions useful, which the author was unable to do. So we do not discuss this any further.

References

[BLR] Bilu, Y. F.; Lev, V.F.; Ruzsa, I.Z.: *Rectification principles in additive number theory.* Discrete Comput. Geom. **19** (1998), no. 3, Special Issue, 343–353.

[C] Cohen, P.J.: *On a conjecture of Littlewood and idempotent measures.* Amer. J. Math. **82** 1960 191–212.

[D1] Davenport, H.: *A Historical Note.* J. London Math. Soc. 1947 s1-22: 100-101.

[D2] Davenport, H.: *On the Addition of Residue Classes.* J. London Math. Soc. 1935 s1-10: 30-32

[DHP] Deshouillers, J.; Hennecart, F.; Plagne, A *On small sumsets in* $(\mathbb{Z}/2\mathbb{Z})^n$. Combinatorica **24** (2004), no. 1, 53–68.

[E] Erdós, P.: *Extremal problems in number theory.* Proceedings of the Symp. Pure Math. VIII AMS (1965), 181189.

[F] Freiman, G.A.: *Foundations of a structural theory of set addition.* Transations of Mathmatical Monographs 37. American Mathematical Society, Providence, R.I., 1973.

[F2] Freiman, G.A.: *Inverse problems of additive number theory. VII. The addition of finite sets. IV. The method of trigonometric sums.* Izv. Vys. U?ebn. Zaved. Matematika 1962 no. **6** (31), 131–144.

[F3] Freiman, G.A.: *?????? ? ???????? ?????? ?????????? ?????? ?????* (Groups and Inverse Problems in Additive Number Theory, in Russian).

[G] Green, B.J.: *The number of squares and* $B_h[g]$ *sets.* Acta Arith. **100** (2001), no. 4, 365–390.

[GR] Green, B.J; Ruzsa, I.Z.: *Sets with small sumset and rectification.* Bull. London Math. Soc. **38** (2006), no. 1, 43–52.

[GS1] Green, B.J.; Sanders, T. *Boolean functions with small spectral norm.* Geom. Funct. Anal. **18** (2008), no. 1, 144–162.

[GS2] Green, B.J; Sanders, T.: *A quantitative version of the idempotent theorem in harmonic analysis.* Ann. of Math. (2) **168** (2008), no. 3, 1025–1054.

[GT] Green, B.J; Tao, T.: *Freiman's theorem in finite fields via extremal set theory.* Combin. Probab. Comput. **18** (2009), no. 3, 335–355.

[HR] Hamidoune, Y.O.; Rdseth, .J.: *An inverse theorem mod p.* Acta Arith. **92** (2000), no. 3, 251–262.

[HSZ] Hamidoune, Y.O.; Serra, O.; Zmor, G.: *On the critical pair theory in abelian groups: beyond Chowla's theorem.* Combinatorica **28** (2008), no. 4, 441–467.

[K] Kemperman, J.H.B.: *On small sumsets in an abelian group.* Acta Math. **103** 1960 63–88.

[L1] Lev, V.F.: *Distribution of points on arcs.* Integers **5** (2005), no. 2, A11, 6 pp.

[L2] Lev, V.F.: *More on points and arcs.* Combinatorial number theory, 347–350, de Gruyter, Berlin, 2007.

[LS] Lev, V.F.; Smeliansky, P.Y.: *On addition of two distinct sets of integers.* Acta Arith. **70** (1995), no. 1, 85–91.

[N] Nathanson, M.B.: *Additive number theory. Inverse problems and the geometry of sumsets.* Graduate Texts in Mathematics, **165**. Springer-Verlag, New York, 1996.

[R1] Rudin, W.: *Fourier Analysis on Groups.* Tracts in Pure Appl. Math., vol. 12, Wiley, New York, 1962.

[R2] Rdseth, .J.: *On the addition of residue classes mod p.* Monatsh. Math. **121** (1996), no. 1-2, 139–143.

[R3] Rdseth, .J.: *On Freiman's 2.4-Theorem.* Skr. K. Nor. Vidensk. Selsk. 2006, no. **4**, 11–18.

[S1] Saeki, S.: *On norms of idempotent measures.* Proc. Amer. Math. Soc. **19** 1968 600–602.

[S2] Saeki, S. *On norms of idempotent measures. II.* Proc. Amer. Math. Soc. **19** 1968 367–371.

[TV] Tao, T.; Vu, V. *Additive combinatorics.* Cambridge Studies in Advanced Mathematics, **105**. Cambridge University Press, Cambridge, 2006.

[V1] Vosper, A.G.: *The critical pairs of subsets of a group of prime order.* J. London Math. Soc. **31** (1956), 200–205.

[V2] Vosper, A.G.: *Addendum to "The critical pairs of subsets of a group of prime order".* J. London Math. Soc. **31** (1956), 280–282.

Department of Mathematics, University of Bristol, Bristol BS8 1TW, England

E-mail address: mazlh@bris.ac.uk