ADDITIVE COMBINATORICS

BEN GREEN

Contents

0.1. **Overview.** These are the notes for the 2024-25 version of the course C3.10 *Additive Combinatorics*, taught at the University of Oxford.

Additive combinatorics is the study of additive questions about finite sets of integers.

We will begin by proving a famous theorem of Roth: every set of integers with positive density contains three distinct elements in arithmetic progression. This proof uses some basic ideas from Fourier analysis, which we will develop from scratch. Then, we will turn to the corresponding question in the group $(\mathbf{Z}/3\mathbf{Z})^n$, where much stronger bounds are known using algebraic methods.

Next we will look at the structure of finite sets A of integers which are almost closed under addition in the sense that their sumset $A + A := \{a_1 + a_2 : a_1, a_2 \in A\}$ is relatively small. The highlight here is Freiman's theorem, which states that any such set has a precise combinatorial structure known as a generalised progression. The proof once again uses some Fourier analysis as well as a host of other ingredients such as the geometry of numbers, which we will develop from first principles.

After that, we will turn to the corresponding question in vector spaces over finite fields. We will introduce entropy methods and describe how they may be used to prove a rather precise description of sets with small sumset in this setting.

Finally, we will look at instances of the sum-product phenomenon, which says that it is impossible for a finite set of integers to be simultaneously additively- and multiplicatively structured. This section draws from a particularly rich set of other mathematical areas, including graph theory, geometry and analysis, as well as previous sections of the course. Nonetheless, prerequisites will be minimal and we will develop what we need from scratch.

A particular aim of the course will be to give a taster of the very large number of different methods which have been brought to bear on these topics: Fourier analysis, algebraic methods, methods from information theory, graph theory and geometric combinatorics.

0.2. **Synopsis.** *Arithmetic progressions.* Basic properties of Fourier transforms. Roth's theorem that every subset of $\{1, \dots, N\}$ of size at least δN contains three elements in arithmetic progression, provided N is sufficiently large in terms of δ. The Croot-Lev-Pach method and strong bounds for arithmetic progressions in $(\mathbf{Z}/3\mathbf{Z})^n$.

Sumsets and Freiman's theorem. Basic sumset estimates. Additive energy and its relation to sumsets: statement (but not proof) of the Balog-Szemerédi-Gowers theorem. Bohr sets and Bogolyubov's theorem. Minkowski's second theorem (statement only). Freiman's theorem on sets with small doubling constant. Freiman's lemma on the dimension of sets with small doubling.

Entropy methods and polynomial Freiman-Ruzsa. Basic notions of entropy and entropy analogues of sumset inequalities. The fibring inequality for entropy doubling. Marton's conjecture in characteristic 2. Deduction of the weak polynomial Freiman-Ruzsa conjecture over $\mathbf{Z}$.

Sum-product theorems. The crossing number inequality for graphs. The Szemerédi-Trotter theorem on point-line incidences, and application to prove that either $|A + A|$ or $|A \cdot A|$ has size at least $c|A|^{5/4}$. Proof of Bourgain and Chang's result that either the m-fold sumset $A + A + \cdots + A$ or the m-fold product set $A \cdot A \cdots A$ has size at least $|A|^{f(m)}$, where $f(m) \to \infty$.

If time allows the course will conclude with a brief non-examinable discussion of Gowers's work on Szemerédi's theorem for progressions of length 4 and longer, which ties together several earlier strands in the course.

0.3. **Further reading.** M. Nathanson's two books [6, 7] have historically been a significant inspiration for the choice of topics in this course. They are a little out of date now, but [7] is still a useful resource for the topics in Sections 3 and 4.

The book of T. Tao and V. Vu [8] is similarly influential and also very useful, though again this book does not cover the more recent developments.

Probably the best external resource for this course is the very new book [9] by Yufei Zhao, which I highly recommend.

The material in Sections 5 to 7 is for the most part very recent. The notes are essentially self-contained, but the reader may also want to consult the original papers, particularly [3, 4].

In Section 5, we will state basic properties of entropy, but will not give the proofs. The proofs of all these statements may be found in the notes for the Oxford undergraduate course B8.4: Information Theory (Chapter 1). A very succinct resource for this material is [1, Section 14.6].

0.4. **Notation.** *Asymptotic notation.* Throughout the course we will be using *asymptotic notation.* This is vital in handling the many inequalities and rough estimates we will encounter. Here is a summary of the notation we will see. We suggest the reader not worry too much about this now; we will gain plenty of practice with this notation. See also the first question on Sheet 0.

- $A \ll B$ means that there is an absolute constant $C > 0$ such that $|A| \leqslant CB$. In this notation, A and B will typically be variable quantities, depending on some other parameter. For example, $x + 1 \ll x$ for $x \geqslant 1$, because $|x + 1| \leqslant 2x$ in this range. It is important to note that the constant C may be different in different instances of the notation.
- $A = O(B)$ means the same thing.
- $A \ll B$ is the same as $B \gg A$.

- $O(A)$ means some quantity bounded in magnitude by CA for some absolute constant $C > 0$. In particular, $O(1)$ simply means a quantity bounded by an absolute positive constant. For example, $\frac{5x}{1+x} = O(1)$ for $x \geqslant 0$.

We shall adopt the very standard notation

$$e(t) := e^{2\pi i t}.$$

One may think of this either as a function on $\mathbf{R}$, periodic with period 1, or as a function on $\mathbf{R}/\mathbf{Z}$; we shall not be careful in making the distinction.

For $\theta \in \mathbf{R}/\mathbf{Z}$ we write $\|\theta\|_{\mathbf{R}/\mathbf{Z}}$ for the distance of θ from 0. Thus, for example, $\|2/3\|_{\mathbf{R}/\mathbf{Z}} = 1/3$.

Write $[N] := \{1, \dots, N\}$. Often, we will state that N is 'sufficiently large' by which we mean larger than some absolute constant that could be specified if desired (but we will generally not do so).

Quantities. In understanding additive and analytic number theory, it is important to develop a robust intuitive feeling for the rough size of certain quantities. For example, you should try and become familiar with facts such as

$$\log^{10} X \ll e^{\sqrt{\log X}} \ll X^{0.01}.$$

1. Roth's theorem on progressions of length 3

In this section our aim is to prove the following theorem of Roth from 1953. In doing so, we introduce a key tool: the Fourier transform.

Theorem 1.1 (Roth's theorem)**.** *Let N be sufficiently large. There is an absolute constant C such that any subset $A \subset [N]$ with cardinality at least $CN/\log\log N$ contains a nontrivial three-term arithmetic progression* (*that is to say, a triple $x, x+d, x+2d$ with $d \neq 0$*)*.*

Note, in particular, that $1/\log\log N$ is eventually smaller than any fixed positive constant.

1.1. The density increment strategy.

Roth's theorem proceeds via the so-called *density increment strategy,* and the key proposition which drives this is the following.

Proposition 1.2. *Let N be sufficiently large. Suppose that $0 < \alpha < 1$ and that $N \geqslant (8/\alpha)^{10}$. Suppose that $P \subset \mathbf{Z}$ is an arithmetic progression of length N and that $A \subset P$ is a set with cardinality at least αN. Then one of the following two alternatives holds:*

(i) *A contains a nontrivial 3-term progression;*

(ii) *There is an arithmetic progression P' of length $N' \geqslant N^{1/5}$ such that, writing $A' := A \cap P'$ and $\alpha' := |A'|/|P'|$, we have $\alpha' \geqslant \alpha + \frac{\alpha^2}{112}$.*

Theorem 1.1 follows by iterating this proposition.

Proof of Theorem 1.1, assuming Proposition 1.2. In this proof c, C, C' denote absolute constants with $0 < c < 1 < C, C'$. Set $P_0 := \{1, \dots, N\}$ and let us suppose that we have a set $A \subset P_0$ with $|A| = \alpha N$ and containing no nontrivial 3-term progression. Then we attempt to use Proposition 1.2 repeatedly to obtain a sequence $P_0, P_1, P_2, \dots$ of progressions together with sets $A_i := A \cap P_i$. The length of P_i will be $N_i \geqslant N^{(1/5)^i}$ and the densities $\alpha_i := |A_i|/|P_i|$ will satisfy $\alpha_{i+1} > \alpha_i + c\alpha_i^2$. (Here we could take $c = \frac{1}{112}$.)

Now this iteration cannot last too long: after C/α steps the density has already doubled, after a further $C/2\alpha$ steps it has doubled again, and so on. Since no set can have density greater than one, there can be no more than $2C/\alpha$ steps in total. We conclude that our applications of Proposition 1.2 must have been invalid, which means that either N_i is not sufficiently large, or $N_i < (8/\alpha_i)^{10}$. Either way, since $\alpha_i \geqslant \alpha$, we have $N_i < C\alpha^{-C}$ for some absolute C. Since

$$N_i > N^{(1/5)^i} \geqslant N^{(1/5)^{2C/\alpha}},$$

we infer the bound

$$N^{(1/5)^{2C/\alpha}} \leqslant C\alpha^{-C}.$$

Rearranging gives

$$\log\log N \leqslant \log\log(C\alpha^{-C}) + \frac{2C}{\alpha}\log 5 \leqslant \frac{C'}{\alpha},$$

which immediately gives the claimed bound. □

Remark. The most important parameter by far is the number of times we performed the iteration, which was roughly $O(1/\alpha)$.

1.2. **Fourier transform on Z.** Let $f : \mathbf{Z} \to \mathbf{C}$ be a compactly-supported function (that is, $f(n) = 0$ outside of some finite interval). Then we define the Fourier transform $\widehat{f}(\theta)$ by

$$\widehat{f}(\theta) := \sum_n f(n)e(-n\theta).$$

Since f is compactly-supported, there is no issue of convergence. A crucial fact we will need is the Parseval identity.

Lemma 1.3. *We have*

$$\sum_n f(n)\overline{g(n)} = \int_{\mathbf{R}/\mathbf{Z}} \widehat{f}(\theta)\overline{\widehat{g}(\theta)}d\theta.$$

Proof. This is an easy check using the definitions, as well as the fact that

$$\int_{\mathbf{R}/\mathbf{Z}} e(m\theta)d\theta = \int_0^1 e(m\theta)d\theta = \begin{cases} 1 & m = 0 \\ 0 & m \in \mathbf{Z}\setminus\{0\}, \end{cases} \tag{1.1}$$

that is to say the orthogonality relation for characters. □

Remark. Taking $f = g$ gives

$$\sum_n |f(n)|^2 = \int_{\mathbf{R}/\mathbf{Z}} |\widehat{f}(\theta)|^2 d\theta.$$

1.3. **A large Fourier coefficient.** We turn now to the details of the density increment strategy. We begin with a very simple observation, which is that we may assume without loss of generality that $P = [N] = \{1, \dots, N\}$. We may always reduce to this case by an affine rescaling.

We will first establish the following alternative version of Proposition 1.2, in which the conclusion of part (ii) is different, asserting the existence of a large Fourier coefficient of the function

$$f_A := 1_A - \alpha 1_{[N]},$$

the so-called *balanced function* of A. In the next section, we will show that a large Fourier coefficient implies a density increment as in the original formulation of Proposition 1.2.

Lemma 1.4. *Suppose that $0 < \alpha < 1$ and that $N \geqslant 4/\alpha^2$. Suppose that $A \subset [N]$ is a set with cardinality at least αN. Then one of the following two alternatives holds:*

(i) *A contains a nontrivial 3-term progression;*

(ii) *The balanced function f_A has a large Fourier coefficient: specifically, there is some $\theta \in \mathbf{R}/\mathbf{Z}$ such that $|\widehat{f_A}(\theta)| \geqslant \alpha^2 N/28$.*

Proof. If $f_1, f_2, f_3 : \mathbf{Z} \to \mathbf{R}$ are three finitely-supported functions then we introduce the operator

$$T(f_1, f_2, f_3) := \sum_{x,d} f_1(x)f_2(x+d)f_3(x+2d).$$

This counts the number of 3-term progressions weighted by the functions f_i. In particular,

$$T(1_A, 1_A, 1_A) = \#\{\text{number of 3-term progressions in } A\}. \tag{1.2}$$

Note carefully that this count includes "trivial" progressions with $d = 0$. However, A has precisely αN trivial progressions, so if option (i) does not hold then

$$T(1_A, 1_A, 1_A) = \alpha N \leqslant \alpha^3 N^2/4. \tag{1.3}$$

For the inequality on the right we used the assumption that $N \geqslant 4/\alpha^2$.

Note that T is a trilinear operator. Thus we may write $1_A = f_A + \alpha 1_{[N]}$ and expand $T(1_A, 1_A, 1_A)$ as a sum of eight terms,

$$T(1_A, 1_A, 1_A) = \alpha^3 T(1_{[N]}, 1_{[N]}, 1_{[N]}) + \cdots + T(f_A, f_A, f_A). \tag{1.4}$$

Each of the seven "error terms" denoted by the ellipsis $\cdots$ contains at least one copy of f_A. Let us look at the first term $\alpha^3 T(1_{[N]}, 1_{[N]}, 1_{[N]})$. It is quite simple to evaluate this exactly: the number of (x, d) with $x, x+d, x+2d \in [N]$ is precisely the number of pairs $(n_1, n_2) \in [N] \times [N]$ with n_1, n_2 having the same parity, since we then have, uniquely, $x = n_1$ and $d = \frac{1}{2}(n_2 - n_1)$, and $x + d$ automatically lies in $[N]$. This is $N^2/2$ if N is even, and $(N^2+1)/2$ if N is odd, thus at least $N^2/2$ in all cases. Thus

$$\alpha^3 T(1_{[N]}, 1_{[N]}, 1_{[N]}) \geqslant \alpha^3 N^2/2.$$

It follows that if option (i) does not hold (and hence we have (1.3)) then the sum of the seven error terms in (1.4) is at least $\alpha^3 N^2/4$. Thus one of these terms is at least $\alpha^3 N^2/28$, that is to say

$$|T(f_1, f_2, f_3)| \geqslant \alpha^3 N^2/28, \tag{1.5}$$

where each f_i is either $\alpha 1_{[N]}$ or f_A, and at least one of them is f_A.

Now we come to the key idea: there is a formula for $T(f_1, f_2, f_3)$ in terms of the Fourier transform:

$$T(f_1, f_2, f_3) = \int_{\mathbf{R}/\mathbf{Z}} \widehat{f_1}(\theta)\widehat{f_2}(-2\theta)\widehat{f_3}(\theta) d\theta. \tag{1.6}$$

Once written down, it is very easy to check this by substituting the definition of the Fourier transforms on the right-hand side and using the orthogonality relations (1.1).

Thus if (1.5) holds then

$$\Big| \int_{\mathbf{R}/\mathbf{Z}} \widehat{f_1}(\theta)\widehat{f_2}(-2\theta)\widehat{f_3}(\theta) d\theta \Big| \geqslant \alpha^3 N^2/28. \tag{1.7}$$

Suppose that $f_3 = f_A$; the analysis of other possibilities is very similar. Then

$$\sup_{\theta \in \mathbf{R}/\mathbf{Z}} |\widehat{f_A}(\theta)| \int_{\mathbf{R}/\mathbf{Z}} |\widehat{f_1}(\theta)||\widehat{f_2}(-2\theta)| d\theta \geqslant \alpha^3 N^2/28.$$

By the Cauchy-Schwarz inequality,

$$\sup_{\theta \in \mathbf{R}/\mathbf{Z}} |\widehat{f_A}(\theta)| \Big(\int_{\mathbf{R}/\mathbf{Z}} |\widehat{f_1}(\theta)|^2 d\theta \Big)^{1/2} \Big(\int_{\mathbf{R}/\mathbf{Z}} |\widehat{f_2}(\theta)| d\theta \Big)^{1/2} \geqslant \alpha^3 N^2/28. \tag{1.8}$$

However, by Parseval's identity we have

$$\int_{\mathbf{R}/\mathbf{Z}} |f_i(\theta)|^2 d\theta = \sum_n |f_i(n)|^2.$$

One may easily check that the RHS is $\alpha^2 N$ if $f_i = \alpha 1_{[N]}$ and $\alpha(1-\alpha)N$ if $f_i = f_A$, and so certainly at most αN in either case. Thus from (1.8) we obtain

$$\sup_{\theta \in \mathbf{R}/\mathbf{Z}} |\widehat{f_A}(\theta)| \geqslant \alpha^2 N/28,$$

which is precisely option (ii) in the proposition. □

1.4. **From a large Fourier coefficient to a density increment.** In this section, we show how option (ii) in Lemma 1.4 (the balanced function f_A has a large Fourier coefficient) may be replaced by option (ii) in Proposition 1.2 (a density increment on a progression). The crucial technical ingredient is the following.

Here, if $F : \mathbf{Z} \to \mathbf{C}$ is a function and $S \subset \mathbf{Z}$ a finite set, we write $\operatorname{diam}_S(F) := \sup_{x,x' \in S} |F(x) - F(x')|$.

Lemma 1.5. *Suppose that N is sufficiently large. Suppose that $\theta \in \mathbf{R}/\mathbf{Z}$. Then we may partition $[N]$ into progressions P_i, each of length at least $N^{1/5}$, such that $\operatorname{diam}_{P_i}(e(\theta x)) \leqslant N^{-1/5}$ for all i.*

Proof. Let $Q := \lfloor N^{1/2} \rfloor$. By a well-known application of the pigeonhole principle due to Dirichlet, there is some positive $d \leqslant Q$ such that $\|d\theta\|_{\mathbf{R}/\mathbf{Z}} \leqslant 1/Q$. (Consider $\theta, 2\theta, \cdots, Q\theta$ as elements of $\mathbf{R}/\mathbf{Z}$; some two of these, say $j_1\theta$ and $j_2\theta$, lie within $1/Q$ of one another. Take $d := |j_1 - j_2|$.)

If P is any progression with common difference d and length $\leqslant 3N^{1/5}$ then, by the triangle inequality,

$$\operatorname{diam}_P(e(\theta x)) \leqslant 3N^{1/5}|e(\theta d) - 1| \leqslant 20N^{1/5}/Q < N^{-1/5},$$

where here we used the inequality

$$|e(t) - 1| = 2|\sin \pi t| \leqslant 2\pi \|t\|_{\mathbf{R}/\mathbf{Z}}.$$

Now observe that $[N]$ can be partitioned into progressions P_i with common difference d and lengths in the range $[N^{1/5}, 3N^{1/5}]$. To do this, first partition $[N]$ into progressions of common difference d, each of length $\sim N/d \gg N^{1/2}$. Then proceed along each such progression from left to right, partitioning into progressions of length $\lceil N^{1/5} \rceil$ until we have a leftover progression of length $\leqslant N^{1/5}$. Amalgamate this with the preceding one. □

The following result, together with Lemma 1.4, immediately implies Proposition 1.2, and hence completes the proof of Roth's theorem.

Lemma 1.6. *Suppose that $|\widehat{f_A}(\theta)| \geqslant \alpha^2 N/28$, that $N \geqslant (8/\alpha)^{10}$, and let $[N] = \bigcup_i P_i$ be a partition as above. Then there is some i such that $|A \cap P_i| \geqslant (\alpha + \frac{\alpha^2}{112})|P_i|$.*

Proof. Since the P_i partition $[N]$, we obviously have

$$\sum_i \Big| \sum_{x \in P_i} f_A(x) e(-\theta x) \Big| \geqslant \frac{\alpha^2}{28} N.$$

By the triangle inequality and the bound $|f_A(x)| \leqslant 1$, the left-hand side is at most

$$\begin{aligned} \sum_i \Big| \sum_{x \in P_i} f_A(x) \Big| + \sum_i |P_i| \operatorname{diam}_{P_i}(e(\theta x)) &\leqslant \sum_i \Big| \sum_{x \in P_i} f_A(x) \Big| + N^{4/5} \\ &\leqslant \sum_i \Big| \sum_{x \in P_i} f_A(x) \Big| + \frac{\alpha^2}{56} N, \end{aligned}$$

the last step following from our assumption on N. It follows that

$$\sum_i \Big| \sum_{x \in P_i} f_A(x) \Big| \geqslant \frac{\alpha^2}{56} N.$$

Since $\sum_{x \in [N]} f_A(x) = 0$, we have

$$\sum_i \Big(\Big| \sum_{x \in P_i} f_A(x) \Big| + \sum_{x \in P_i} f_A(x) \Big) \geqslant \frac{\alpha^2}{56} N = \frac{\alpha^2}{56} \sum_i |P_i|,$$

so there must be some i such that

$$\Big|\sum_{x \in P_i} f_A(x)\Big| + \sum_{x \in P_i} f_A(x) \geqslant \frac{\alpha^2}{56}|P_i|,$$

which implies that

$$\sum_{x \in P_i} f_A(x) \geqslant \frac{\alpha^2}{112}|P_i|,$$

or in other words that

$$|A \cap P_i| \geqslant \big(\alpha + \frac{\alpha^2}{112}\big)|P_i|.$$

This concludes the proof. □

2. Progressions in finite fields and the polynomial method

In this section we consider questions about 3-term arithmetic progressions in a so-called finite field model. We will focus on the specific case of progressions in the group $\mathbf{F}_3^n$, where $\mathbf{F}_3 \cong \mathbf{Z}/3\mathbf{Z}$ is the finite field of order 3, and n is a large integer, though similar techniques may be used in $\mathbf{F}_p^n$ for any odd prime p. Characteristic 3 is quite attractive because a 3-term progression is then just a triple of points with $x + y + z = 0$ (since $z = -2z$ in characteristic 3).

Fourier analytic methods may be used to attack this problem. There are three questions on Example Sheet 1 in which the details of this are worked out (or you are invited to work them out).

The purpose of this chapter is to give a fairly recent (2016) very different proof, which gives a dramatically better bound for the problem in this setting. The following is a theorem of Ellenberg and Gijswijt, who adapted a breakthrough of Croot, Lev and Pach.

Theorem 2.1. *Suppose that $A \subset \mathbf{F}_3^n$ is a set containing no three elements in arithmetic progression. Then $|A| \ll (3-\delta)^n$ for some positive constant δ.*

We will follow a presentation of the argument due to Tao. In what follows, write $\mathbf{F} = \mathbf{F}_3$.

Definition 2.2. Let $A \subset \mathbf{F}^n$, and suppose that $f : A \times A \times A \to \mathbf{F}$ is a function. Then the *slice rank* $\mathrm{sr}(f)$ of f is the smallest r for which f may be written as the sum of r functions of the form $g(x)h(y,z)$, $g(y)h(x,z)$ or $g(z)h(x,y)$.

Lemma 2.3. *Suppose that f as above has support exactly equal to the diagonal $x = y = z$; that is, $f(x,y,z)$ is nonzero if and only if $x = y = z$. Then $\mathrm{sr}(F) = |A|$.*

Before giving the proof, we isolate a (well-known) linear-algebraic lemma.

Lemma 2.4. *Any k-dimensional subspace of $\mathbf{F}^n$ contains a point with at least k nonzero coordinates in the standard basis.*

Proof. Let V be the subspace. Write down a $k \times n$ matrix whose rows are a basis for V, in the standard coordinate system on $\mathbf{F}^n$. Since $\dim V = k$, the row rank of this matrix is k. Therefore its column rank is also k. By permuting coordinates, we may suppose without loss of generality that the $k \times k$ submatrix consisting of the first k columns has full rank. The point $(1,1,\dots,1,0,0,\dots,0)$ (with k 1s and $(n-k)$ 0s) then lies in V. □

We may now prove Lemma 2.3.

Proof of Lemma 2.3. We first note that *any* f (regardless of support properties) has slice rank at most $|A|$. This is quite obvious by writing

$$f(x,y,z) = \sum_{a \in A} \mathbf{1}_{a=x} f(a,y,z).$$

The main content is therefore the bound in the other direction. Suppose that $f(x,y,z)$ is a sum of r_1 functions of the form $g_{1,i}(x)h_{1,i}(y,z)$, r_2 functions of the form $g_{2,i}(y)h_{2,i}(x,z)$ and r_3 functions of the form $g_{3,i}(z)h_{3,i}(x,y)$, with $r_1+r_2+r_3=r$.

We claim that there is a function $\phi : A \to \mathbf{F}$ with

$$\sum_x \phi(x)g_{3,i}(x) = 0, \qquad i = 1,2,\dots,r_3, \tag{2.1}$$

and for which $\phi(x)$ is nonzero for at least $|A|-r_3$ values of $x \in A$. Indeed, the space of ϕ satisfying (2.1) is a vector space of dimension at least $|A|-r_3$, and so we may apply Lemma 2.3.

Now consider

$$\tilde{f}(x,y) := \sum_{a\in A} f(x,y,a)\phi(a).$$

Since

$$\sum_{a\in A} g_{3,i}(a)h_{3,i}(x,y)\phi(a) = 0,$$

whilst

$$\sum_{a\in A} g_{2,i}(y)h_{2,i}(x,a)\phi(a) = g_{2,i}(y)\Big(\sum_{a\in A} h_{2,i}(x,a)\phi(a)\Big)$$

has the form $\alpha(x)\beta(y)$, as does

$$\sum_{a\in A} g_{1,i}(x)h_{1,i}(y,a)\phi(a),$$

we see that $\tilde{f}(x,y)$ is a sum of r_1+r_2 functions of the form $\alpha(x)\beta(y)$, that is to say the rank of the matrix $M := (\tilde{f}(x,y))_{x,y\in A}$ is at most r_1+r_2.

Note, however, that M is diagonal by the assumptions on f, that is to say $\tilde{f}(x,y)=0$ if $x \neq y$. Moreover, the diagonal entry $\tilde{f}(x,x)$ is $f(x,x,x)\phi(x)$, which vanishes if and only if $\phi(x)$ does; thus there are at least $|A|-r_3$ nonzero diagonal entries, and so the rank of M is at least $|A|-r_3$.

Combining these two inequalities, we see that

$$r_1+r_2 \geqslant |A|-r_3,$$

or in other words $r \geqslant |A|$, as desired. □

Now suppose that $A \subset \mathbf{F}_3^n$ is a set with no nontrivial solution to $x+y+z=0$ (that is, no 3-term progression). Then the function $f : A\times A\times A \to \mathbf{F}$ defined by $f(x,y,z)=1$ if $x+y+z=0$, and 0 otherwise, has exactly the property in Lemma 2.3. Thus we conclude from Lemma 2.3 that $\mathrm{sr}(f)=|A|$. To complete the proof of Theorem 2.1, it therefore suffices to prove that $\mathrm{sr}(f) \leqslant (3-\delta)^n$. In fact we will prove that the similar function $F : \mathbf{F}^n\times\mathbf{F}^n\times\mathbf{F}^n \to \mathbf{F}$ defined by $F(x,y,z)=1$ if $x+y+z=0$, and 0 otherwise (that is, the extension of f from $A\times A\times A$ to $\mathbf{F}^n\times\mathbf{F}^n\times\mathbf{F}^n$) has slice rank $\leqslant (3-\delta)^n$, which is a stronger result since clearly $\mathrm{sr}(F)\geqslant \mathrm{sr}(f)$. The following proposition establishes a sharp form of this claim.

Proposition 2.5. *Let F be as above. Then $\mathrm{sr}(F) \leqslant 3(1+t+t^2)^n t^{-2n/3}$ for any $t \in (0,1]$.*

Proof. The key idea is to observe that

$$F(x,y,z) = \prod_{i=1}^n \big(1-(x_i+y_i+z_i)^2\big),$$

using here the fact that $1-u^2=0$ unless $u=0$ (in $\mathbf{F}=\mathbf{F}_3$). Expanding out, this is a polynomial of degree $2n$ in the $3n$ variables x_i,y_i,z_i. For each monomial, at least one of the total x-degree, total y-degree or total z-degree is $\leqslant 2n/3$. The sum of the terms with total x-degree at most $2n/3$ may be written as a sum of terms $m(x)h_m(y,z)$, where m ranges over all monomials $m(x)=x_1^{i_1}\cdots x_n^{i_n}$

of total degree $\leqslant 2n/3$, and $h_m(y,z)$ is some function. Therefore, $\mathrm{sr}(F)$ is at most 3 times the number of monomials m of degree $\leqslant 2n/3$.

Expand the product

$$\prod_{i=1}^{n}(1+x_i+x_i^2)$$

as a sum of monomials, and set $x_1 = \cdots = x_n = t$, for some $t \in [0,1]$. Each monomial in the x_i (of degree at most 2 in each x_i) appears, and the ones with degree d contribute t^d each. Thus, the number of monomials of degree d is at most $t^{-d}(1+t+t^2)^n$, and so by the preceding discussion we see that

$$\mathrm{sr}(F) \leqslant 3t^{-2n/3}(1+t+t^2)^n,$$

as required. □

Combining Proposition 2.5 with the remarks immediately preceding it, we see that if $A \subset \mathbf{F}_3^n$ has no 3-term progression then $|A| \leqslant 3t^{-2n/3}(1+t+t^2)^n$. Here, we are free to choose $t \in (0,1]$.

Setting $t = 1-\varepsilon$ and expanding to first order, we see that $t^{-2/3}(1+t+t^2) = 3 - \varepsilon + O(\varepsilon^2)$, so by taking ε sufficiently small we obtain the claimed bound $|A| \ll (3-\delta)^n$.

Alternatively, one may compute the optimal value of t using calculus; this is Sheet 1, Q4.

3. Sumset inequalities

In this section we explore the notion of adding sets. There is a huge literature on this topic, from which we isolate a few key results. All of the results we shall state are valid for finite subsets of arbitrary abelian groups, and for brevity it is usual to call these "additive sets". When we are talking about more than one additive set, we assume they are all subsets of the same group. The particular abelian group in question will normally be clear from context (though often it does not matter). In fact, many of the results (but not all) remain true without the assumption of commutativity, but we shall not cover that topic in this course.

3.1. **Basic notation and definitions.** Let A, B be additive sets (this both A and B are finite subsets of some abelian group). Then we write

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

and

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

These definitions extend in an obvious way to more than two summands, for example

$$A_1 + \cdots + A_k := \{a_1 + \cdots + a_k : a_i \in A_i\}.$$

If $A_1 = \cdots = A_k = A$ then we usually write kA for $A_1 + \cdots + A_k$. In particular, $2A = A + A$. We also write, e.g. $2A - 2A$ for $\{a_1 + a_2 - a_3 - a_4 : a_1, \dots, a_4 \in A\}$.

3.2. **Ruzsa's triangle inequality and covering lemma.** In this section we prove two elegant results of Ruzsa about the size of sumsets. They are surprisingly useful despite their apparent simplicity.

Lemma 3.1 (Ruzsa triangle inequality)**.** *Suppose that U, V, W are finite additive sets. Then*

$$|V - W||U| \leqslant |V - U||U - W|.$$

Proof. We will define a map $\phi : (V-W)\times U \to (V-U)\times(U-W)$, and prove that it is an injection, which implies the result. Given $d \in V-W$ select a pair $v_d \in V, w_d \in W$ for which $d = v_d - w_d$ (there may be more than one such pair, but for each d we make a definite choice). Then define

$$\phi(d,u) = (v_d - u, u - w_d)$$

for each $d \in V - W$ and $u \in U$. To prove that ϕ is an injection, suppose that $(x,y) \in \mathrm{im}(\phi) \subset (V-U)\times(U-W)$. If $\phi(d,u) = (x,y)$ then $x+y = (v_d-u)+(u-w_d) = v_d - w_d = d$, and therefore we can determine d and hence v_d and w_d from (x,y). And we also determine u as $u = -x + v_d$ $(= y - w_d)$. □

Remark. If we define

$$d(U,V) := \log \frac{|U-V|}{|U|^{1/2}|V|^{1/2}}$$

then the Ruzsa triangle inequality may be written

$$d(V,W) \leqslant d(U,V) + d(U,W).$$

This explains the term "triangle inequality". Note that, although the triangle inequality is satisfied, d is not a true distance. This is because $d(U,V) = 0$ neither implies, nor is implied by, $U = V$.

Lemma 3.2 (Ruzsa's covering lemma)**.** *Suppose that A and B are finite additive sets and that $|A+B| \leqslant K|A|$. Then B may be covered by k translates of $A-A$, for some $k \leqslant K$. That is, there is a set X, $|X| \leqslant K$, such that*

$$B \subset (A-A) + X.$$

Proof. Choose $X \subset B$ maximal so that $\{A + x : x \in X\}$ are disjoint. The union of these sets contains exactly $|A||X|$ elements, and all of these elements lie in $A+B$. Therefore $|X| \leqslant K$. Now, if $b \in B$ then $A+b$ intersects $A+x$ for some $x \in X$, because of the maximality of X, and so $b \in A - A + x$. Hence, $B \subset (A-A)+X$. □

3.3. **Petridis's inequality.** In this section and the next we develop inequalities controlling the size of sums of three or more sets. A beautiful way to do this was discovered surprisingly recently by Petridis. His result is stated as Corollary 3.6 below. We give an elegant rephrasing of his proof which was given by Tao on the blog of Tim Gowers.

Let B be a set in some abelian group G. Let K be a real number, and consider the function ϕ on subsets of G defined by

$$\phi(A) := |A+B| - K|A|. \tag{3.1}$$

Lemma 3.3. *ϕ is* submodular*, that is to say it satisfies*

$$\phi(A \cup A') + \phi(A\cap A') \leqslant \phi(A) + \phi(A').$$

Proof. Write $\sigma(A) := A+B$. Observe that

$$\sigma(A\cup A') = \sigma(A)\cup\sigma(A'),$$

and that

$$\sigma(A\cap A') \subseteq \sigma(A)\cap\sigma(A').$$

Therefore

$$\begin{aligned} |\sigma(A\cup A')| = |\sigma(A)\cup\sigma(A')| &= |\sigma(A)| + |\sigma(A')| - |\sigma(A)\cap\sigma(A')| \\ &\leqslant |\sigma(A)| + |\sigma(A')| - |\sigma(A\cap A')|, \end{aligned}$$

that is to say $|\sigma|$ satisfies the submodularity property

$$|\sigma(A\cup A')| + |\sigma(A\cap A')| \leqslant |\sigma(A)| + |\sigma(A')|.$$

Since the function $|A|$ satisfies

$$|A \cup A'| + |A \cap A'| = |A| + |A'|,$$

the result follows immediately, since $\phi(A) = |\sigma(A)| - K|A|$. □

Lemma 3.4. *Let ϕ be any submodular function. Suppose that $A_1, \dots, A_n$ are sets with the following property: $\phi(A_i) = 0$, and $\phi(Z_i) \geqslant 0$ for every subset $Z_i \subseteq A_i$. Then $\phi\big(\bigcup_{i=1}^n A_i\big) \leqslant 0$.*

Proof. By the assumptions and submodularity, for any i and for any set S, we have

$$\phi(A_i \cup S) \leqslant \phi(A_i \cup S) + \phi(A_i \cap S) \leqslant \phi(A_i) + \phi(S) = \phi(S).$$

The result then follows immediately by induction on n. □

Proposition 3.5 (Petridis)**.** *Let A, B be sets in some abelian group. Suppose that $|A+B| = K|A|$ and that $|Z+B| \geqslant K|Z|$ for all $Z \subseteq A$. Then, for any further set S in the group, $|A+B+S| \leqslant K|A+S|$.*

Proof. Apply Lemma 3.4 with the particular function ϕ defined in (3.1) above. Take the A_i to be the translates $A + s$ of A by elements of s. It is easy to check that the hypotheses of Lemma 3.4 hold. Observe that $\bigcup_{i=1}^n A_i = A + S$, and so the Lemma implies that $\phi(A+S) \leqslant 0$, or in other words $|A+B+S| \leqslant K|A+S|$. □

It is convenient to apply Petridis's inequality in the following form.

Corollary 3.6. *Let A, B be sets in some abelian group. Suppose that $|A+B| \leqslant K|A|$. Let $X \subseteq A$ be a non-empty set for which the ratio $|X+B|/|X|$ is minimal. Then for any further set S we have*

$$|S+X+B| \leqslant K|S+X|.$$

Proof. Apply Proposition 3.5 with A replaced by X. □

3.4. **The Plünnecke–Ruzsa inequality.** The most widely applicable result about higher-order sumsets is the Plünnecke–Ruzsa inequality.

Theorem 3.7 (Plünnecke–Ruzsa)**.** *Suppose that A and B are additive sets with $|A+B| \leqslant K|A|$. Let $k, \ell \geqslant 0$ be integers. Then $|kB - \ell B| \leqslant K^{k+\ell}|A|$.*

The original proof was quite long and involved a fair amount of machinery from graph theory. Nowadays, it can be deduced quickly from Petridis's inequality.

Lemma 3.8. *Suppose that A and B are finite additive sets for which $|A+B| \leqslant K|A|$. Then there exists $X \subset A$ for which $|X+kB| \leqslant K^k|X|$.*

Proof. Let X be the subset of A for which the ratio $|X+B|/|X|$ is minimal. By Petridis's inequality (Corollary 3.6) with $S = (k-1)B$, we have

$$|X+kB| = |X+(k-1)B+B| \leqslant K|X+(k-1)B|.$$

The result then follows by induction on k. □

Proof of Theorem 3.7. . Suppose that A and B are finite additive sets for which $|A+B| \leqslant K|A|$. By Ruzsa's Triangle Inequality with U, V, W replaced by $X, -kB, -\ell B$, respectively, and then Lemma 3.8, we have

$$|kB - \ell B|\;|X| \leqslant |X+kB| \cdot |X+\ell B| \leqslant K^{k+\ell}|X|^2.$$

Thus, since $X \subset A$, $|kB - \ell B| \leqslant K^{k+\ell}|X| \leqslant K^{k+\ell}|A|$. □

Let us record a corollary of this and Lemma 3.1 concerning the relationship between sums and differences.

Corollary 3.9. *Suppose that A, B are additive sets. Then*

$$|A \mp B| \leqslant \frac{|A \pm B|^3}{|A||B|}.$$

Proof. We handle the case with a minus sign on the left and a plus sign on the right; the other case follows immediately from this by switching B to $-B$. Set $K := |A+B|/|A|$. First apply Theorem 3.7 with $k = 2$ and $\ell = 0$ to obtain $|B+B| \leqslant K^2|A|$. Now, apply Lemma 3.1 with $U = B$, $V = -B$ and $W = -A$ to get $|A-B||B| \leqslant |B+B||A+B|$. Combining these two estimates gives the result. □

3.5. **Additive energy and Balog–Szemerédi-Gowers.** In this section we introduce the concept of additive energy, which is closely related to the notion of sumset and arises naturally in applications (such as in Section 9).

We have already seen the notion of an additive set having small doubling. The next definition introduces some notation for this, and also introduces a kind of bipartite variant of the concept which applies to pairs of sets.

Definition 3.10. Let A be an additive set. Then we define the *doubling constant*

$$\sigma[A] := \frac{|A+A|}{|A|}.$$

If A, B are two additive sets, we write

$$\sigma[A, B] := \frac{|A+B|}{|A|^{1/2}|B|^{1/2}}.$$

Note that $\sigma[A] = \sigma[A, A]$, so one may think of the former as a shorthand for the latter. The notion of a set having small doubling is somehow "combinatorial" in that it refers to the *size* of $|A+A|$ and does not take account, for example, of the number of representations. The notion has some serious shortcomings, for example being highly sensitive to small changes to A.

In this subsection we explore the related notion of additive energy, which is more "analytic", more robust to small perturbations, and often arises in nature.

Definition 3.11. Let A be an additive set. Then we define the *additive energy* $E(A)$ to be the number of *additive quadruples* in A, that is to say quadruples $(a_1, a_2, a_3, a_4) \in A^4$ such that $a_1 + a_2 = a_3 + a_4$. We define the normalised additive energy $\omega[A]$ to be $E(A)/|A|^3$. More generally if A, B are two additive sets, we write

$$E(A, B) := \#\{(a, b, a', b') \in A \times B \times A \times B : a + b = a' + b'\}$$

and

$$\omega[A, B] := |A|^{-3/2}|B|^{-3/2}E(A, B).$$

Note that $0 \leqslant \omega(A) \leqslant 1$: the upper bound here follows from the fact that three elements of an additive quadruple uniquely determine the fourth. More generally, $0 \leqslant \omega[A, B] \leqslant 1$. This follows from the fact that $E(A, B) \leqslant \min(|A|^2|B|, |A||B|^2)$, since any three elements of a quadruple (a, b, a', b') satisfying $a + b = a' + b'$ determine the fourth. However, $\min(|A|^2|B|, |A||B|^2) \leqslant |A|^{3/2}|B|^{3/2}$.

Proposition 3.12. *We have $\sigma[A, B]\omega[A, B] \geqslant 1$. In particular, if the doubling constant of a pair A, B of additive sets is at most K, their normalised additive energy is at least $1/K$. In particular, specialising to the case $A = B$, we have $\sigma[A]\omega[A] \geqslant 1$.*

Proof. For $x \in A + B$ write $r(x)$ for the number of pairs $(a, b) \in A \times B$ with $a + b = x$. Then

$$\sum r(x) = |A||B|,$$

whilst

$$\sum r(x)^2 = E(A, B).$$

Moreover, $r(x)$ is supported (that is, is nonzero) on $A + B$. Thus by Cauchy-Schwartz

$$|A|^2|B|^2 = \big(\sum_x r(x)\big)^2 \leqslant |A + B| \sum_x r(x)^2 = |A + B|E(A, B),$$

which rearranges to give the stated inequality. □

The converse to this kind of statement fails dramatically, as the following shows.

Example. Let n be a large even number. Let $A_1 = \{1, \dots, n/2\}$ and let A_2 be some arbitrary set of $n/2$ integers having no additive relation with A_1, for instance $A_2 = \{10^n, 10^{2n}, \dots, 10^{n^2/2}\}$. Set $A = A_1 \cup A_2$, a set of size n. Then

$$E(A) \geqslant E(A_1) \geqslant \frac{1}{10}n^3,$$

but

$$|A + A| \geqslant |A_2 + A_2| \geqslant \frac{1}{8}n^2,$$

since the sums of pairs in A_2 are distinct apart from the relations $x + y = y + x$. Thus $\omega[A] \geqslant \frac{1}{10}$, but $\sigma[A]$ grows linearly in n.

The Balog-Szemerédi-Gowers theorem is a remarkable result which nonetheless salvages a kind of partial converse to Proposition 3.12. We state a bipartite and a single set version of the result. Recall that by convention C is an absolute constant which can change from line to line (but could be written in explicitly wherever it occurs, if desired, with a bit of work).

Theorem 3.13 (Balog-Szemerédi-Gowers)**.** *We have the following statements.*

(i) *Suppose that A, B are additive sets and that $\omega[A, B] \geqslant 1/K$. Then there are sets $A' \subseteq A$, $B' \subseteq B$ with $|A'| \geqslant K^{-C}|A|$, $|B'| \geqslant K^{-C}|B|$ such that $\sigma[A', B'] \ll K^C$.*

(ii) *Suppose that A is an additive set and that $\omega[A] \geqslant 1/K$. Then there is a set $A' \subset A$ with $|A'| \gg K^{-C}|A|$ such that $\sigma[A'] \ll K^C$.*

The proof of the Balog–Szemerédi–Gowers course is not examinable, but I will go over it in lectures at the end if time allows. For notes, see Appendix A. The value of C we obtain there is quite reasonable in principle, but we will not be too concerned with computing an exact value.

4. Freiman's theorem

This section contains one of the highlights of the course, which is a fairly complete (at least qualitatively) answer to the question of what sets with small sumset look like. Let us begin with a little context for the question.

Recall that if A is a set of integers then

$$A + A := \{a_1 + a_2 : a_1, a_2 \in A\}.$$

Suppose A has size n. How big is $A + A$? Trivially, it has size at most $\frac{1}{2}n(n + 1)$, that being the number of pairs (a_1, a_2), with (a_1, a_2) and (a_2, a_1) counted the same. On the other hand, it has size at least $2n - 1$. Writing $a_1 < \cdots < a_n$ for the elements of A, we have

$$a_1 + a_1 < a_1 + a_2 < \cdots < a_1 + a_n < a_2 + a_n < \cdots < a_n + a_n,$$

a listing of $2n-1$ distinct elements of A. Equality can occur in both bounds. For example if $A = \{1, 2, \dots, 2^{n-1}\}$ then all the sums $a_1 + a_2$ are distinct (except for the trivial relations $a_1 + a_2 = a_2 + a_1$). If $A = \{1, \dots, n\}$ then $A + A = \{2, \dots, 2n\}$, a set of size $2n - 1$.

We say that A has *doubling at most* K if

$$\sigma[A] = \frac{|A+A|}{|A|} \leqslant K.$$

Typically, we will have in mind that K is fixed (say $K = 10$) and $n = |A|$ is very large. The basic question to be considered in this section is that of what we can say about the structure of A if $\sigma[A] \leqslant K$, for some small K.

4.1. **Generalised progressions and Freiman's theorem.** Before stating the main result, let us give some progressively more complicated motivating examples.

Example (Progression). Let A be any arithmetic progression of length n. Then $|A + A| = 2n - 1$.

Example (Subsets of progressions). Let P be a progression of length Cn, and let $A \subset P$ be an arbitrary set of size n. Then $|A + A| \leqslant 2Cn$.

Example (2-dimensional progression). Suppose that $L_1L_2 = n$, and consider a set A of the form

$$A := \{x_0 + \ell_1 x_1 + \ell_2 x_2 : 0 \leqslant \ell_1 < L_1, 0 \leqslant \ell_2 < L_2\}.$$

If the x_i are suitably widely spaced, the elements described here are all distinct and $|A| = n$. In this case we say that A is *proper*. We have

$$A + A = \{2x_0 + \ell_1' x_1 + \ell_2' x_2 : 0 \leqslant \ell_1' < 2L_1 - 1, 0 \leqslant \ell_2' < 2L_2 - 1\},$$

and so certainly

$$|A + A| \leqslant 4|A|.$$

Example (d-dimensional progression). The same as above, but with d parameters $L_1, \dots, L_d$: thus

$$A = \{x_0 + l_1 x_1 + \cdots + l_d x_d : 0 \leqslant l_i < L_i\}. \tag{4.1}$$

Now, if A is proper, we have $|A + A| \leqslant 2^d |A|$.

Example (Subsets of multidimensional progressions). Let P be a proper d-dimensional progression of size Cn. Let $A \subset P$ be an arbitrary set of size n. Then

$$|A + A| \leqslant |P + P| \leqslant 2^d |P| = 2^d Cn.$$

The final example gives a somewhat large class of sets with doubling constant at most K (pick any parameters d, C with $2^d C \leqslant K$).

Freiman's theorem is the result that the above examples are the only ones.

Theorem 4.1 (Freiman)**.** *Suppose that $A \subset \mathbf{Z}$ is a finite set with $|A + A| \leqslant K|A|$. Then A is contained in a generalised progression P of dimension $\ll_K 1$ and size $\ll_K |A|$.*

The *size* of a generalised progression as in (4.1) is defined to be $L_1 \cdots L_d$. This is at least the cardinality of the progression, but is strictly bigger than it if the progression fails to be proper.

Freiman's theorem states that A is contained in a proper progression of dimension at most $d(K)$ and size at most $C(K)|A|$, where $d(), C()$ are functions of K only. In this course we will not be concerned with bounds, but the argument we give leads to a bound for $d(K)$ that is exponential in K, and a bound for $C(K)$ that is doubly exponential in K. This is quite far from the truth; in fact, it does not require a vast amount of further effort to remove an exponential from both of these bounds, but we will not do so here.

Many other refinements are possible, but again we will not cover them here. For example, one can insist that P be proper if desired.

4.2. **Freiman homomorphisms.** In his remarkably insightful 1966 book, Freiman made an attempt to treat additive number theory by analogy with the way Klein treated geometry: as well as sets $A, B, \cdots$ of integers, one should study maps between them and, most particularly, properties invariant under natural types of map. This was doubtless regarded as somewhat eccentric at the time, but the notion of Freiman homomorphism is now quite important in additive combinatorics.

Definition 4.2. Suppose that $s \geqslant 2$ is an integer. Suppose that A, B are additive sets. Then we say that a map $\phi : A \to B$ is a Freiman s-homomorphism if we have

$$\phi(a_1) + \cdots + \phi(a_s) = \phi(a_1') + \cdots + \phi(a_s')$$

whenever

$$a_1 + \cdots + a_s = a_1' + \cdots + a_s'.$$

It is obvious that any group homomorphism restricts to a Freiman homomorphism (of arbitrary order) on any subset. However, the notion is much more general. For example, any map whatsoever from $A = \{1, 10, 100, 1000\}$ to another additive set is a Freiman 2-homomorphism, simply because A has no nontrivial relations of the form $a_1 + a_2 = a_1' + a_2'$.

The map ϕ is said to be a Freiman s-isomorphism if it has an inverse ϕ^{-1} which is also a Freiman s-homomorphism. We caution that, contrary to what is often expected in more algebraic situations, a one-to-one Freiman homomorphism need not be a Freiman isomorphism. For example, the obvious map

$$\phi : \{0, 1\}^n \to (\mathbf{Z}/2\mathbf{Z})^n$$

is a Freiman homomorphism of all orders (it is induced from the natural group homomorphism $\mathbf{Z}^n \to (\mathbf{Z}/2\mathbf{Z})^n$). However, it is not a Freiman 2-isomorphism as $(\mathbf{Z}/2\mathbf{Z})^n$ contains a great many more additive relations than $\{0, 1\}^n$.

The following lemma records some basic facts about Freiman isomorphisms.

Lemma 4.3. *Suppose that A, B, C are additive sets. Let $s \geqslant 2$ be an integer. Then we have the following.*

(i) *Suppose that $\phi : A \to B$ and $\psi : B \to C$ are Freiman s-homomorphisms. Then so is the composition $\psi \circ \phi$.*

(ii) *Suppose that $\phi : A \to B$ is a Freiman s-homomorphism. Then it is also a Freiman s'-homomorphism for every s' satisfying $2 \leqslant s' \leqslant s$.*

(iii) *Suppose that $\phi : A \to B$ is a Freiman s-homomorphism and let $k, l \geqslant 0$ be integers. Then ϕ induces a Freiman s'-homomorphism $\tilde{\phi} : kA - lA \to kB - lB$, for any integer $s' \leqslant s/(k+l)$.*

(iv) *The above three statements also hold with "homo" replaced by "iso" throughout.*

(v) *Suppose that P is a generalised progression and that $\phi : P \to B$ is a Freiman 2-homomorphism. Then $\phi(P)$ is a generalised progression of the same dimension. If ϕ is a Freiman 2-isomorphism, and if P is proper, then so is $\phi(P)$.*

(vi) *Let $\pi_m : \mathbf{Z} \to \mathbf{Z}/m\mathbf{Z}$ be the natural map. Then π_m is a Freiman s-isomorphism when restricted to $(t, t + \frac{m}{s}] \cap \mathbf{Z}$, for any $t \in \mathbf{R}$.*

Proof. The first four parts of this are very straightforward once one has understood the definitions, and we will not go over them carefully in lectures. Perhaps (iii) requires some further comment: one should define $\tilde{\phi} : kA - lA \to kB - lB$ by

$$\tilde{\phi}(a_1 + \cdots + a_k - a_1' - \cdots - a_l') = \phi(a_1) + \cdots + \phi(a_k) - \phi(a_1') - \cdots - \phi(a_l').$$

One must then check that this is well-defined and is a Freiman homomorphism of the order claimed.

To prove (v), let $\phi : P \to \phi(P)$ be a Freiman 2-homomorphism. Suppose that $P = \{x_0 + l_1x_1 + \cdots + l_dx_d : 0 \leqslant l_i < L_i\}$. Set $y_0 = \phi(x_0)$, and define $y_1, \dots, y_d$ by $y_0 + y_i = \phi(x_0 + x_i)$ for $i = 1, \dots, d$; we claim that $\phi(x_0 + l_1x_1 + \cdots + l_dx_d) = y_0 + l_1y_1 + \cdots + l_dy_d$ for all $l_1, \dots, l_d$ satisfying

$0 \leqslant l_i < L_i$. This may be established by induction on $l_1 + \dots + l_d$, noting that we have defined the y_i in such a way that it holds whenever $l_1 + \dots + l_d = 0$ or 1. To obtain the statement for $(l_1, \dots, l_d) = (1, 1, 0, \dots, 0)$, for example, one may use the relation

$$x_0 + (x_0 + x_1 + x_2) = (x_0 + x_1) + (x_0 + x_2)$$

to conclude that

$$\phi(x_0) + \phi(x_0 + x_1 + x_2) = \phi(x_0 + x_1) + \phi(x_0 + x_2)$$

and hence that $\phi(x_0 + x_1 + x_2) = y_0 + y_1 + y_2$, as required.

Finally, we comment on (vi). Since π_m is a group homomorphism, it is also a Freiman homomorphism. Its restriction to any interval of length at most m is a bijection. Suppose that $x_1, \dots, x_s, x_1', \dots, x_s'$ satisfy $t < x_i, x_i' \leqslant t + \frac{m}{s}$ and that $\pi_m(x_1) + \dots + \pi_m(x_s) = \pi_m(x_1') + \dots + \pi_m(x_s')$, that is to say $x_1 + \dots + x_s = x_1' + \dots + x_s' (\operatorname{mod} m)$. Then, since $|x_1 + \dots + x_s - x_1' - \dots - x_s'| < m$, we must have $x_1 + \dots + x_s = x_1' + \dots + x_s'$. □

4.3. **Ruzsa's model lemma.** In this section we prove a remarkable lemma of Imre Ruzsa. It asserts that a subset of $\mathbf{Z}$ with small doubling has a large piece which is Freiman isomorphic to a dense subset of a cyclic group $\mathbf{Z}/m\mathbf{Z}$. In that setting the tools of harmonic analysis become much more powerful, unlike for arbitrary subsets of $\mathbf{Z}$ (even those of small doubling) which could well be highly "spread out". Here is Ruzsa's lemma.

Proposition 4.4 (Ruzsa model lemma)**.** *Suppose that $A \subset \mathbf{Z}$ is a finite set and that $s \geqslant 2$ is an integer. Let $m \geqslant |sA - sA|$ be an integer. Then there is a set $A' \subset A$ with $|A'| \geqslant |A|/s$ which is Freiman s-isomorphic to a subset of $\mathbf{Z}/m\mathbf{Z}$.*

Proof. By translating A if necessary, we may assume that A consists of positive integers. Let q be a very large prime number. For $\lambda \in (\mathbf{Z}/q\mathbf{Z})^\times$, consider the composition $\phi_\lambda := \gamma \circ \beta \circ \alpha_\lambda$ of maps

$$\mathbf{Z} \xrightarrow{\alpha_\lambda} \mathbf{Z}/q\mathbf{Z} \xrightarrow{\beta} \{1, \dots, q\} \xrightarrow{\gamma} \mathbf{Z}/m\mathbf{Z}$$

where

- α_λ is reduction mod q followed by multiplication by λ;
- β inverts the reduction mod q map;
- γ is the reduction mod m map.

Now α_λ and γ are Freiman homomorphisms of all orders. The map β is not, but by Lemma 4.3 (vi), it is a Freiman s-homomorphism on the reduction mod q of any interval $I_j := \{n \in \mathbf{Z} : \frac{jq}{s} < n \leqslant \frac{(j+1)q}{s}\}$. Since s such intervals (with $j = 0, 1, \dots, q-1$) cover $\{1, \dots, q\}$, it follows by the pigeonhole principle that for every λ there is a $j = j(\lambda)$ such that the set

$$A_\lambda := \{a \in A : \alpha_\lambda(a) \in I_{j(\lambda)} (\operatorname{mod} q)\}$$

has size at least $|A|/s$. By construction, ϕ_λ is a Freiman s-homomorphism when restricted to A_λ.

Everything we have said so far holds for arbitrary λ. To complete the proof, we now show that there exists some λ such that ϕ_λ is invertible, and its inverse is a Freiman s-homomorphism. If this fails for some λ then this means that there is

$$d = a_1 + \dots + a_s - a_1' - \dots - a_s' \neq 0$$

such that

$$\phi_\lambda(a_1) + \dots + \phi_\lambda(a_s) = \phi_\lambda(a_1') + \dots + \phi_\lambda(a_s'). \tag{4.2}$$

Here, d and the $a_i, a_i' \in A_\lambda$ depend on λ.

Write

$$x := \sum_{i=1}^{s} \beta(\alpha_\lambda(a_i)) - \sum_{i=1}^{s} \beta(\alpha_\lambda(a_i')).$$

Then if (4.2) holds we have $\gamma(x) = 0$, that is to say $x \equiv 0 (\mathrm{mod}\, m)$.

Without loss of generality (switching the a_i and the a_i' if necessary) we may assume that $x \geqslant 0$. Also, since the a_i, a_i' lie in A_λ, it follows that $x \in s(I_{j(\lambda)} - I_{j(\lambda)}) \subset (-q, q)$. Therefore $0 \leqslant x < q$.

Now by construction, x and λd are congruent modulo q. It therefore follows that

$$x = \psi(\lambda d),$$

where $\psi(n)$ is the unique integer in $\{0, 1, \dots, q-1\}$ congruent to n modulo q.

From now on we indicate the dependence of d on λ explicitly. To summarise, we have shown the following. If $\phi_\lambda|_{A_\lambda}$ is not a Freiman s-isomorphism, then there must be some $d_\lambda \in (sA - sA) \setminus \{0\}$ such that

$$\psi(\lambda d_\lambda) \equiv 0 (\mathrm{mod}\, m).$$

To get a contradiction, Let us fix $d \in (sA - sA) \setminus \{0\}$ and ask about values of λ for which $d = d_\lambda$: lacking imagination, we call them "bad for d". If the prime q is chosen big enough then it will not divide any element of $(sA - sA) \setminus \{0\}$, so d is coprime to q.

As λ ranges over $(\mathbf{Z}/q\mathbf{Z})^\times$, λd covers $(\mathbf{Z}/q\mathbf{Z})^\times$ uniformly, and hence the set $\{\psi(\lambda d) : \lambda \in (\mathbf{Z}/q\mathbf{Z})^\times\}$ coincides with $\{1, \dots, q-1\}$. The number of elements y in this interval for which $y \equiv 0 (\mathrm{mod}\, m)$ is at most $(q-1)/m$. Since each d lies in the set $(sA - sA) \setminus \{0\}$, it follows that the number of λ which are bad for *some* d is at most

$$\frac{q-1}{m}\big(|sA - sA| - 1\big) < q - 1,$$

the inequality being a consequence of the assumption that $m \geqslant |sA - sA|$.

This is a contradiction, since *every* λ is bad for some d (namely d_λ). □

In our proof of Freiman's theorem, we will use the following corollary.

Corollary 4.5. *Suppose that $A \subset \mathbf{Z}$ is a finite set with doubling constant K. Then there is a prime $q \leqslant 2K^{16}|A|$ and a subset $A' \subset A$ with $|A'| \geqslant |A|/8$ such that A' is Freiman 8-isomorphic to a subset of $\mathbf{Z}/q\mathbf{Z}$.*

Proof. By the Plünnecke–Ruzsa inequality, Theorem 3.7, we have $|8A - 8A| \leqslant K^{16}|A|$. Now by Bertrand's postulate there is a prime q satisfying $|8A - 8A| \leqslant q \leqslant 2|8A - 8A|$. This prime of course satisfies the bound $q \leqslant 2K^{16}|A|$, and by the Ruzsa model lemma there is a subset $A' \subset A$ with $|A'| \geqslant |A|/8$ which is Freiman 8-isomorphic to a subset of $\mathbf{Z}/q\mathbf{Z}$. □

4.4. **Bogolyubov's lemma.** Ruzsa's model lemma (or, more accurately, Corollary 4.5) allows us to switch attention from a set $A \subset \mathbf{Z}$ with small doubling to a dense subset of a cyclic group $\mathbf{Z}/q\mathbf{Z}$. We now prove a lemma about the structure of such sets.

Definition 4.6. Suppose that $R = \{r_1, \dots, r_k\}$ is a set of nonzero elements of $\mathbf{Z}/q\mathbf{Z}$ and that $\varepsilon > 0$ is a parameter. Then we define the *Bohr set* $B(R, \varepsilon)$ with frequency set R and width ε by

$$B(R, \varepsilon) := \{x \in \mathbf{Z}/q\mathbf{Z} : \Big\| \frac{r_i x}{q} \Big\|_{\mathbf{R}/\mathbf{Z}} \leqslant \varepsilon \text{ for } i = 1, 2, \dots, k\}.$$

The parameter k is said to be the *dimension* of the Bohr set.

Proposition 4.7 (Bogolyubov's lemma)**.** *Let $S \subset \mathbf{Z}/q\mathbf{Z}$ be a set of size σq. Then $2S - 2S$ contains a Bohr set of dimension at most $4/\sigma^2$ and width at least $\frac{1}{10}$.*

In the proof, we will use the discrete Fourier transform, specifically the Fourier transform on $\mathbf{Z}/q\mathbf{Z}$. (The Fourier transform can in fact be developed on any locally compact abelian group, and in this way the Fourier transform on $\mathbf{Z}$ which featured in Section 1, the discrete Fourier transform on $\mathbf{Z}/q\mathbf{Z}$, and the Fourier transform on $(\mathbf{Z}/3\mathbf{Z})^n$ discussed on Example Sheet 1 may be considered as special cases of the same general concept.) Here are the relevant definitions and basic properties.

Definition 4.8. Let $f : \mathbf{Z}/q\mathbf{Z} \to \mathbf{C}$ be a function. Then for $r \in \mathbf{Z}/q\mathbf{Z}$ we define the (discrete) Fourier transform

$$\widehat{f}(r) := \frac{1}{q} \sum_{x \in \mathbf{Z}/q\mathbf{Z}} f(x) e(-rx/q).$$

Proposition 4.9. *In the following proposition, $f, g : \mathbf{Z}/q\mathbf{Z} \to \mathbf{C}$ are two functions.*

(i) *We have the inversion formula*

$$f(x) = \sum_{r \in \mathbf{Z}/q\mathbf{Z}} \widehat{f}(r) e(rx/q).$$

(ii) *We have the Parseval identity*

$$\frac{1}{q} \sum_{x \in \mathbf{Z}/q\mathbf{Z}} f(x) \overline{g(x)} = \sum_{r \in \mathbf{Z}/q\mathbf{Z}} \widehat{f}(r) \overline{\widehat{g}(r)}.$$

(iii) *If the convolution $f * g : \mathbf{Z}/q\mathbf{Z} \to \mathbf{C}$ is defined by*

$$(f * g)(x) := \frac{1}{q} \sum_{y \in \mathbf{Z}/q\mathbf{Z}} f(y) g(x - y)$$

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

Proof. Once again, all of this is an easy check using the definitions, as well as the fact that

$$\sum_r e(rx/q) = \begin{cases} q & x = 0 \\ 0 & x \in (\mathbf{Z}/q\mathbf{Z}) \setminus \{0\}. \end{cases} \qquad \square$$

Remark. Taking $f = g$ in the Parseval identity gives

$$\frac{1}{q} \sum_{x \in \mathbf{Z}/q\mathbf{Z}} |f(x)|^2 = \sum_{r \in \mathbf{Z}/q\mathbf{Z}} |\widehat{f}(r)|^2.$$

It is worth pausing to consider what convolution "does". If f and g are functions supported on sets $A, B \subseteq \mathbf{Z}/q\mathbf{Z}$ respectively (for instance, we could have $f = 1_A$ and $g = 1_B$) then $f * g$ is supported on $A + B$. Moreover, $f * g$ has a nice Fourier transform, which can be very convenient for further analysis. Note carefully that $1_A * 1_B$ is not the same thing as 1_{A+B}; the latter function puts equal weight on every element of $A + B$, whereas the former weights elements x according to the number of representations as $a + b$ with $a \in A, b \in B$.

Now we turn to the proof of Bogolyubov's lemma, Proposition 4.7.

Proof of Proposition 4.7. Consider the function $f := 1_S * 1_S * 1_{-S} * 1_{-S}$. This is supported on $2S - 2S$, that is to say if $f(x) > 0$ then $x \in 2S - 2S$. Note also that $\widehat{1_{-S}}(r) = \overline{\widehat{1_S}(r)}$, and so $\widehat{f}(r) = |\widehat{1_S}(r)|^4$. By the Fourier inversion formula and the fact that f is real, we have

$$f(x) = \sum_{r \in \mathbf{Z}/q\mathbf{Z}} |\widehat{1_S}(r)|^4 e(rx/q) = \sum_{r \in \mathbf{Z}/q\mathbf{Z}} |\widehat{1_S}(r)|^4 \cos(2\pi rx/q). \tag{4.3}$$

Let R be the set of all $r \neq 0$ for which $|\widehat{1_S}(r)| \geqslant \sigma^{3/2}/2$. By Parseval's identity we have

$$|R| \frac{\sigma^3}{4} \leqslant \sum_{r \in R} |\widehat{1_S}(r)|^2 \leqslant \sum_{r \in \mathbf{Z}/q\mathbf{Z}} |\widehat{1_S}(r)|^2 = \frac{1}{q} \sum_{x \in \mathbf{Z}/q\mathbf{Z}} 1_S(x)^2 = \sigma,$$

and so

$$|R| \leqslant 4/\sigma^2. \tag{4.4}$$

We claim that $B(R, \frac{1}{10}) \subset 2S - 2S$, to which end it suffices to show that $f(x) > 0$ for $x \in B(R, \frac{1}{10})$. To do this, we will use the formula (4.3). We split the sum over r into three pieces: the term $r = 0$, the terms with $r \in R$, and all other terms. Clearly

$$|\widehat{1_S}(0)|^4 = \sigma^4.$$

If $r \in R$ then $\cos(2\pi rx/q) \geqslant 0$, so the sum of these terms is nonnegative. Finally,

$$\sum_{r \notin R \cup \{0\}} |\widehat{1_S}(r)|^4 \cos(2\pi rx/q) \geqslant - \sum_{r \notin R \cup \{0\}} |\widehat{1_S}(r)|^4 \geqslant -\frac{\sigma^3}{4} \sum_r |\widehat{1_S}(r)|^2 = -\frac{\sigma^4}{4},$$

the last step being a further application of Parseval's identity. Combining all of this we obtain

$$f(x) \geqslant \sigma^4 + 0 - \frac{\sigma^4}{4} > 0,$$

as required. □

4.5. **Generalised progressions in Bohr sets.** It is by no means obvious what has been gained in proving Proposition 4.7. The answer is that a Bohr set $B(R, \varepsilon)$ has a great deal of structure, in particular containing a large generalised progression. The key proposition is as follows.

Proposition 4.10. *Let $R \subset \mathbf{Z}/q\mathbf{Z}$ be a set of size k, not containing zero. Let $0 < \varepsilon < \frac{1}{2}$. Then the Bohr set $B(R, \varepsilon)$ contains a proper generalised progression of dimension k and cardinality at least $(\varepsilon/k)^k q$.*

In the proof, we will rely on a result from the geometry of numbers, Minkowski's second theorem. This is stated as Proposition 4.11 below. The proof is not examinable, but it is given in Appendix C. To even state the theorem, we need some terminology.

A *lattice* $\Lambda \subset \mathbf{R}^d$ is a discrete and cocompact subgroup of $\mathbf{R}^d$. It is a theorem that every lattice is of the form $\mathbf{Z}v_1 \oplus \mathbf{Z}v_2 \oplus \cdots \oplus \mathbf{Z}v_d$ for linearly independent $v_1, \dots, v_d$, which are then called an *integral basis* for Λ. The set $\mathcal{F} := \{x_1 v_1 + \cdots + x_d v_d : 0 \leqslant x_i < 1\}$ is then called a fundamental region for Λ; note that translates of it by Λ precisely cover $\mathbf{R}^d$. Note that the v_i (and hence $\mathcal{F}$) are not uniquely determined by Λ, but it turns out that the volume of $\mathcal{F}$ is. The determinant $\det(\Lambda)$ is the volume of a fundamental region of Λ.

The statement of Minkowski's Second Theorem also involves a centrally symmetric convex body $K \subset \mathbf{R}^d$. This means a set which is convex (meaning that if $x, y \in K$ then $\lambda x + (1-\lambda)y \in K$ for all $\lambda \in [0,1]$) and centrally symmetric, which means that if $x \in K$ then $-x \in K$.

The geometry of numbers is, to an extent, the study of how lattices Λ interact with convex bodies K.

Suppose we have a lattice Λ and a convex body K. We define the *successive minima* $\lambda_1, \dots, \lambda_d$ of K with respect to Λ as follows: λ_j is the infimum of those λ for which the dilate λK contains j linearly independent elements of Λ. If K is compact then $\lambda_j K$ itself contains j linearly independent elements of Λ. (For each $\varepsilon > 0$, $(\lambda_j + \varepsilon)K$ contains such elements. Since these all lie in $(\lambda_j + 1)K$, there are only finitely many choices, and in particular for some sequence of ε tending to zero we may make the same choice. Since K is compact, these elements all lie in $\lambda_j K$.)

Proposition 4.11 (Minkowski's Second Theorem)**.** *We have $\lambda_1 \cdots \lambda_d \operatorname{vol}(K) \leqslant 2^d \det(\Lambda)$.*

We now turn to the proof of Proposition 4.10.

Proof of Proposition 4.10. Let $R = \{r_1, \dots, r_k\}$ and consider the lattice

$$\Lambda = q\mathbf{Z}^k + (r_1, \dots, r_k)\mathbf{Z}.$$

Since q is prime, this may be written as a direct sum

$$q\mathbf{Z}^k \oplus \{0, 1, \dots, q-1\} \cdot (r_1, \dots, r_k).$$

Thus $q\mathbf{Z}^k$ has index q as a subgroup of Λ, and from this and the fact that $\det(q\mathbf{Z}^k) = q^k$ it follows that $\det(\Lambda) = q^{k-1}$ (see Lemma C.1).

Take $K \subset \mathbf{R}^k$ to be the box $\{\mathbf{x} : \|\mathbf{x}\|_\infty \leqslant \varepsilon q\}$. Let $\lambda_1, \dots, \lambda_k$ be the successive minima of K with respect to Λ. Since K is closed, $\lambda_j K$ contains j linearly independent elements of Λ. We may, by choosing each element in turn, select a basis $\mathbf{b}_1, \dots, \mathbf{b}_k$ for $\mathbf{R}^k$ with $\mathbf{b}_j \in \Lambda \cap \lambda_j K$ for all j. (Such a basis is called a *directional basis*; we should caution that, whilst the $\mathbf{b}_j$ are linearly independent elements of Λ, they need not form an integral basis for Λ.) Thus $\mathbf{b}_j \in \Lambda$ and $\|\mathbf{b}_j\|_\infty \leqslant \lambda_j \varepsilon q$. Set $L_j := \lceil 1/\lambda_j k \rceil$ for $j = 1, \dots, k$. Then if $0 \leqslant l_j < L_j$ we have $\|l_j \mathbf{b}_j\|_\infty \leqslant \varepsilon q/k$ and therefore

$$\|l_1 \mathbf{b}_1 + \cdots + l_k \mathbf{b}_k\|_\infty \leqslant \varepsilon q.$$

Now each $\mathbf{b}_i$ lies in Λ and hence is congruent to $x_i(r_1, \dots, r_k) (\operatorname{mod} q)$ for some x_i, $0 \leqslant x_i < q$. Abusing notation slightly, we think of these x_i as lying in $\mathbf{Z}/q\mathbf{Z}$. The preceding observation implies that

$$\Big\| \frac{(l_1 x_1 + \cdots + l_k x_k) r_i}{q} \Big\|_{\mathbf{R}/\mathbf{Z}} \leqslant \varepsilon$$

for each i, or in other words the generalised progression $\{l_1 x_1 + \cdots + l_k x_k : 0 \leqslant l_i < L_i\}$ is contained in the Bohr set $B(R, \varepsilon)$.

It remains to prove a lower bound on the size of this progression and also to establish its properness. The lower bound on the size is easy: it is at least $k^{-k}(\lambda_1 \cdots \lambda_k)^{-1}$ which, by Minkowski's Second Theorem and the fact that $\det(\Lambda) = q^{k-1}$ and $\operatorname{vol}(K) = (2\varepsilon q)^k$, is at least $(\varepsilon/k)^k q$.

To establish the properness, suppose that

$$l_1 x_1 + \cdots + l_k x_k = l_1' x_1 + \cdots + l_k' x_k (\operatorname{mod} q),$$

where $|l_i|, |l_i'| < \lceil 1/k\lambda_i \rceil$. Then the vector

$$\mathbf{b} = (l_1 - l_1')\mathbf{b}_1 + \cdots + (l_k - l_k')\mathbf{b}_k$$

lies in $q\mathbf{Z}^k$ and furthermore

$$\|\mathbf{b}\|_\infty \leqslant \sum_{i=1}^{k} 2 \Big\lfloor \frac{1}{\lambda_i k} \Big\rfloor \|\mathbf{b}_i\|_\infty \leqslant 2\varepsilon q.$$

Since we are assuming that $\varepsilon < 1/2$ it follows that $\mathbf{b} = 0$ and hence, due to the linear independence of the $\mathbf{b}_i$, that $l_i = l_i'$ for all i. Therefore the progression is indeed proper. □

4.6. **Freiman's theorem: conclusion of the proof.** In this section, we conclude the proof of Freiman's theorem.

Proof of Theorem 4.1. By Corollary 4.5, the corollary of Ruzsa's model lemma, there is a prime $q \leqslant 2K^{16}|A|$ and a subset $A' \subset A$ with $|A'| \geqslant |A|/8$ such that A' is Freiman 8-isomorphic to a subset $S \subset \mathbf{Z}/q\mathbf{Z}$. If $\sigma := |S|/q$ then we have $\sigma \geqslant \frac{1}{16}K^{-16}$.

By Bogolyubov's lemma, Proposition 4.7, $2S - 2S$ contains a Bohr set of dimension at most $2^{10}K^{32}$ and width at least $\frac{1}{10}$.

By Proposition 4.10, that Bohr set (and hence $2S - 2S$) contains a proper generalised progression P of dimension at most $K^{O(1)}$ and cardinality at least $\exp(-K^{O(1)})q$. (We could keep track of exact constants, but this becomes a little tedious).

Now A' is Freiman 8-isomorphic to S, and so by Lemma 4.3 (iii), $2A' - 2A'$ is Freiman 2-isomorphic to $2S - 2S$. The inverse of this Freiman isomorphism restricts to a Freiman isomorphism $\phi : P \to \phi(P) \subset 2A' - 2A'$. By Lemma 4.3 (v), $Q = \phi(P)$ is also a proper generalised progression, of the same dimension and size as P. Therefore we have shown that $2A - 2A$ contains a proper generalised progression Q of dimension $K^{O(1)}$ and

$$|Q| \geqslant \exp(-K^{O(1)})|A|. \tag{4.5}$$

To finish the argument, we apply the covering lemma, Lemma 3.2, to the sets Q and A. Since

$$Q + A \subset (2A - 2A) + A = 3A - 2A,$$

the Plünnecke–Ruzsa inequality and (4.5) imply that

$$|Q + A| \leqslant K^5|A| \leqslant \exp(K^{O(1)})|Q|.$$

By Lemma 3.2, there is some set $Y = \{y_1, \dots, y_m\}$,

$$m \leqslant \exp(K^{O(1)}), \tag{4.6}$$

such that $A \subset (Q - Q) + Y$. Suppose that

$$Q = \{x_0 + l_1x_1 + \dots + l_dx_d : 0 \leqslant l_i < L_i\}$$

and that $Y = \{y_1, \dots, y_m\}$. Then

$$(Q - Q) + Y \subset \{\tilde{x}_0 + l_1x_1 + \dots + l_dx_d + l'_1y_1 + \dots + l'_my_m, 0 \leqslant l_i < 2L_i, 0 \leqslant l'_j < 2\} := \tilde{Q}$$

where

$$\tilde{x}_0 = -(L_1x_1 + \dots + L_dx_d).$$

Note that $\tilde{Q}$ is a generalised progression of dimension $d + m$ and that

$$\operatorname{size}(\tilde{Q}) = 2^{d+m}L_1 \cdots L_d = 2^{d+m}|Q| \leqslant 2^{d+m}|2A - 2A| \ll_K |A|,$$

the penultimate step following since $Q \subset 2A - 2A$.

The dominant term in the bound is 2^m, which is double exponential in K. □

4.7. **Freiman's lemma.** We conclude this section with a nice geometric result (not directly related to the earlier results of the section) about small doubling and dimension. It is known as *Freiman's lemma.*

Proposition 4.12 (Freiman's lemma)**.** *Suppose that $A \subset \mathbf{R}^r$ is a finite set, not contained in any affine subspace. Then we have the lower bound*

$$|A + A| \geqslant (r + 1)|A| - \frac{1}{2}r(r + 1).$$

Proof. The set $A + A$ has the same size as the set $m(A) := \frac{1}{2}(A + A)$ of midpoints of line segments of A (note that $A \subseteq m(A)$). Let $F(r, n)$ denote the minimum value of $|m(A)|$ amongst all sets $A \subset \mathbf{R}^r$ which are not contained in an affine subspace and for which $|A| = n$. Consider an extreme point a on the convex hull of A. The set $A' := A \setminus \{a\}$ is either contained in an $(r-1)$-dimensional affine subspace, or it is not. In the former case we clearly have $m(A) \geqslant m(A') + n$, since none of the midpoints of the line segments $[ax]$, $x \in A$, lies in $m(A')$. In the latter case we have $m(A) \geqslant m(A') + r + 1$. Indeed if S is the r-face nearest to a then none of the midpoints of the segments $[ax]$, $x \in S$, lie in $m(A')$, and nor does a.

Both of the cases here are compatible with the inequality

$$F(r, n) \geqslant \min(F(r - 1, n - 1) + n, F(r, n - 1) + r + 1).$$

It follows by induction on $r + n$ (with the base case $r = 0$, $n = 1$ being obvious) that

$$F(r, n) \geqslant (r + 1)n - \frac{1}{2}r(r + 1).$$

The result follows. □

Remark. The argument at the beginning of the section, showing that $|A + A| \geqslant 2|A| - 1$ when A is a set of integers, is essentially a special case of the above proof.

5. Entropy methods

In Sheet 2, Q5 you saw how the methods of Section 3 may be used to prove the following result via a short (but clever) argument of Ruzsa.

Theorem 5.1. *Suppose that $A \subset \mathbf{F}_2^n$ is non-empty with $|A+A| \leqslant K|A|$. Then A is contained in a subgroup $H \leqslant \mathbf{F}_2^n$ of cardinality at most $K^2 2^{K^4}|A|$.*

The main aim of the next few sections is to give the proof of the following more refined result.

Theorem 5.2. *Let $A \subset \mathbf{F}_2^n$ be non-empty with $|A+A| \leqslant K|A|$. Then there exists a subgroup H of $\mathbf{F}_2^n$ with $|H| \leqslant |A|$ such that A is covered by at most $2K^C$ translates of H, for some absolute constant C.*

This result is known as Polynomial Freiman-Ruzsa (PFR); before it was proven, it was known as Marton's Conjecture after Katalin Marton, who first posed the question. The conjecture was proven in 2023 in the paper [3]. We will organise the argument slightly differently (which leads to worse constants, but is arguably easier to understand). The reader may find it helpful to refer to [3] (in particular the appendices) when following this part of the course, though the notes are fairly self-contained.

One of the key ideas in the proof is the introduction of an entropy variant of the notion of sumset. This will be the main focus of the current section.

5.1. Entropy.

We begin with a very brief overview/review of the notion of entropy. For proofs, see (the Oxford course) B8.4: Information Theory (Chapter 1) or [1, Section 14.6].

We will always be working with random variables with finite range. Let X be such a random variable, and write $p_X(x) := \mathbf{P}(X = x)$ for the density function of X. The *range* of X is defined to be the set $\{x : p_X(x) > 0\}$.

We define the entropy $\mathbf{H}[X]$ by

$$\mathbf{H}[X] := \sum_x p_X(x) \log \frac{1}{p_X(x)}.$$

For us, logs will always be natural logs; in theoretical computer science (and in the course B8.4), logs are usually to base 2. It does not make an important difference to the theory.

If X takes values in a set S then

$$\mathbf{H}[X] \leqslant \log |S| \tag{5.1}$$

with equality if and only if X is uniform on S. This is a consequence of Jensen's inequality and the concavity of log. Also, denoting by p_X the density function of X,

$$\mathbf{H}[X] = \sum_x p_X(x) \log \frac{1}{p_X(x)} \geqslant \min_{x : p_X(x) > 0} \log \frac{1}{p_X(x)},$$

and therefore

$$\max_x p_X(x) \geqslant e^{-\mathbf{H}[X]}. \tag{5.2}$$

Since entropy only depends on the values of $p_X(x)$ and not on what the x are, we have

$$\mathbf{H}[X] = \mathbf{H}[\phi(X)] \tag{5.3}$$

whenever ϕ is an injective map on the range of X.

Given a pair (X, Y) of random variables, the *conditional entropy* $\mathbf{H}[X \mid Y]$ is defined by the formula

$$\mathbf{H}[X \mid Y] := \sum_y p_Y(y) \mathbf{H}[X \mid Y = y]$$

where y ranges over the support of p_Y, and $(X \mid Y = y)$ denotes the random variable X conditioned on the event $Y = y$. We have the fundamental *chain rule*

$$\mathbf{H}[X,Y] = \mathbf{H}[X \mid Y] + \mathbf{H}[Y]. \tag{5.4}$$

Here we abbreviate $\mathbf{H}[(X,Y)]$ as $\mathbf{H}[X,Y]$, and will make similar abbreviations regarding other information-theoretic quantities in this paper without further comment; for instance, $\mathbf{H}[(X,Y) \mid (Z,W)]$ becomes $\mathbf{H}[X,Y \mid Z,W]$. Note that (5.4) implies a conditional generalization

$$\mathbf{H}[X,Y \mid Z] = \mathbf{H}[X \mid Y,Z] + \mathbf{H}[Y \mid Z].$$

for all random variables X,Y,Z.

The *mutual information* $\mathbf{I}[X : Y]$ is defined by the formula

$$\begin{aligned}\mathbf{I}[X : Y] &= \mathbf{H}[X] + \mathbf{H}[Y] - \mathbf{H}[X,Y]\\ &= \mathbf{H}[X] - \mathbf{H}[X \mid Y]\\ &= \mathbf{H}[Y] - \mathbf{H}[Y \mid X],\end{aligned}$$

and is non-negative by another application of Jensen's inequality, vanishing precisely when X,Y are independent; in particular

$$\mathbf{H}[X,Y] = \mathbf{H}[X] + \mathbf{H}[Y] \tag{5.5}$$

if and only if X,Y are independent, and

$$\mathbf{H}[X \mid Y] \leqslant \mathbf{H}[X], \qquad \mathbf{H}[X,Y] \leqslant \mathbf{H}[X] + \mathbf{H}[Y] \tag{5.6}$$

always.

Suppose now that (X,Y,Z) is a triple of random variables. Applying (5.6) to $(X \mid Z = z)$ and summing over z (weighted by $p_Z(z)$) gives

$$\mathbf{H}[X \mid Y,Z] \leqslant \mathbf{H}[X \mid Z], \tag{5.7}$$

which is known as *submodularity*. It may equivalently be written as

$$\mathbf{H}[X,Y,Z] + \mathbf{H}[Z] \leqslant \mathbf{H}[X,Z] + \mathbf{H}[Y,Z]. \tag{5.8}$$

The *conditional mutual information* $\mathbf{I}[X : Y \mid Z]$ is defined by

$$\begin{aligned}\mathbf{I}[X : Y \mid Z] &:= \sum_z p_Z(z)\mathbf{I}[(X \mid Z = z) : (Y \mid Z = z)]\\ &= \mathbf{H}[X \mid Z] + \mathbf{H}[Y \mid Z] - \mathbf{H}[X,Y \mid Z].\end{aligned}$$

Submodularity is equivalent to the statement that

$$\mathbf{I}[X : Y \mid Z] \geqslant 0, \tag{5.9}$$

since, as can be seen by expanding,

$$\begin{aligned}\mathbf{I}[X : Y|Z] &= \mathbf{H}[X,Z] + \mathbf{H}[Y,Z] - \mathbf{H}[X,Y,Z] - \mathbf{H}[Z]\\ &= \mathbf{H}[X \mid Z] - \mathbf{H}[X \mid Y,Z].\end{aligned} \tag{5.10}$$

Equality occurs in (5.9) (and hence in (5.8)) if and only if X,Y are conditionally independent relative to Z, which means that the random variables $(X \mid Z = z)$ and $(Y \mid Z = z)$ are independent for every z in the range of Z (that is, for which $p_Z(z) > 0$).

Finally, U_A denotes the uniform distribution on a set A. Note that

$$\mathbf{H}[U_A] = \log |A|. \tag{5.11}$$

5.2. **Entropic Ruzsa distance.** Let G be an abelian group (in much of what follows, we will take $G = \mathbf{F}_2^n$, but for now take G to be arbitrary). Let X, Y be G-valued random variables. Then we define the entropic Ruzsa distance

$$\mathrm{d}[X;Y] := \mathbf{H}[X' - Y'] - \tfrac{1}{2}\mathbf{H}[X] - \tfrac{1}{2}\mathbf{H}[Y],$$

where X', Y' are independent copies of X, Y. (Note that it is convenient to define $\mathrm{d}[X;Y]$ for variables X, Y which may not themselves be independent, including in the extreme case $X = Y$.) It turns out that $\exp(\mathrm{d}[U_A; U_A])$ is a kind of entropic substitute for the doubling constant of the set A (or, more accurately, of $|A - A|/|A|$). For more details on this, see Sheet 3 Q9.

Entropic analogues of Ruzsa distance/doubling constant enjoy the best features of combinatorial sumset notions *and* additive energy at the same time.

For the next several sections we will say 'distance' rather that 'entropic Ruzsa distance' when discussing random variables.

Before moving on to slightly deeper results, we record the fact that distance is non-negative. In fact, we will establish a quantitative version of this which will be useful several times later on.

Lemma 5.3. *Let X, Y be G-valued random variables. Then*

$$\mathrm{d}[X;Y] \geqslant \tfrac{1}{2}|\mathbf{H}[X] - \mathbf{H}[Y]|.$$

Proof. We may assume that X, Y are independent, so that

$$\mathrm{d}[X;Y] = \mathbf{H}[X - Y] - \tfrac{1}{2}\mathbf{H}[X] - \tfrac{1}{2}\mathbf{H}[Y].$$

By (5.6) we have $\mathbf{H}[X - Y] \geqslant \mathbf{H}[X - Y \mid Y]$. On the other hand,

$$\mathbf{H}[X - Y \mid Y] = \mathbf{H}[X - Y, Y] - \mathbf{H}[Y] = \mathbf{H}[X, Y] - \mathbf{H}[Y] = \mathbf{H}[X].$$

Here, in the last step we used independence, and $\mathbf{H}[X - Y, Y] = \mathbf{H}[X, Y]$ by (5.3) since there is an bijection $(X - Y, Y) \mapsto (X, Y)$ induced by $(a, b) \mapsto (a + b, b)$.

Combining these facts gives $\mathbf{H}[X - Y] \geqslant \frac{1}{2}(\mathbf{H}[X] - \mathbf{H}[Y])$, and the corresponding inequality with the roles of X, Y reversed follows similarly. □

The first more serious result is the entropic analogue of the Ruzsa triangle inequality (which we will call 'the triangle inequality' in the next few sections).

Lemma 5.4. *Let A, B, C be G-valued random variables.*

$$\mathrm{d}[A;B] \leqslant \mathrm{d}[A;C] + \mathrm{d}[C;B].$$

Proof. This is equivalent to establishing

$$\mathbf{H}[A - B] \leqslant \mathbf{H}[A - C] + \mathbf{H}[C - B] - \mathbf{H}[C] \tag{5.12}$$

whenever A, B, C are independent. To prove this, the key step is to apply (5.7) in the form

$$\mathbf{H}[B - C \mid A - B] \geqslant \mathbf{H}[B - C \mid A - B, B] = \mathbf{H}[C \mid A, B] = \mathbf{H}[C]$$

where in the last step we used independence. Moreover,

$$\begin{aligned}\mathbf{H}[B - C \mid A - B] &= \mathbf{H}[B - C, A - B] - \mathbf{H}[A - B] \\ &= \mathbf{H}[A - C, B - C] - \mathbf{H}[A - B] \\ &\leqslant \mathbf{H}[A - C] + \mathbf{H}[B - C] - \mathbf{H}[A - B]\end{aligned}$$

by (5.6). Combining these gives (5.12). □

One may observe that, in the above proof, we did not fully use the assumption that A, B, C are independent: all that was used is that C is independent of (A, B). Thus we may upgrade Lemma 5.4 to the statement that

$$\mathbf{H}[A - B] - \tfrac{1}{2}\mathbf{H}[A] - \tfrac{1}{2}\mathbf{H}[B] \leqslant \mathrm{d}[A; C] + \mathrm{d}[C; B], \tag{5.13}$$

where here A, B need not be independent. We will use this inequality in one place in Section 7.

The following inequality of Madiman/Kaimanovich–Vershik is the entropy analogue of Petridis's inequality.

Lemma 5.5. *Let A, B, C be independent G-valued random variables. Then we have*

$$\mathbf{H}[A + B + C] - \mathbf{H}[A + B] \leqslant \mathbf{H}[B + C] - \mathbf{H}[B].$$

Proof. By (5.10) we have

$$\mathbf{I}[A : C \mid A + B + C] = \mathbf{H}[A, A + B + C] + \mathbf{H}[C, A + B + C] - \mathbf{H}[A, C, A + B + C] - \mathbf{H}[A + B + C].$$

However, using (5.5) three times we have $\mathbf{H}[A, A + B + C] = \mathbf{H}[A, B + C] = \mathbf{H}[A] + \mathbf{H}[B + C]$, $\mathbf{H}[C, A + B + C] = \mathbf{H}[C, A + B] = \mathbf{H}[C] + \mathbf{H}[A + B]$ and $\mathbf{H}[A, C, A + B + C] = \mathbf{H}[A, B, C] = \mathbf{H}[A] + \mathbf{H}[B] + \mathbf{H}[C]$.

After a short calculation, we see that the claimed inequality is equivalent to the assertion that $\mathbf{I}[A : C \mid A + B + C] \geqslant 0$, which is an instance of submodularity in the form (5.9). □

The other key advantage of the entropic notions is that they behave well under projections in a way that combinatorial notions do not. The key result here is a result we call the fibring identity.

Here, and in several places in what follows, we need conditioned notions of distance. If (X, Z) and (Y, W) are random variables (where X and Y are G-valued) we define

$$\mathrm{d}[X \mid Z; Y \mid W] := \sum_{z,w} p_Z(z) p_W(w) \mathrm{d}[(X \mid Z = z); (Y \mid W = w)]. \tag{5.14}$$

Alternatively, if $(X', Z'), (Y', W')$ are independent copies of the variables (X, Z), (Y, W),

$$\mathrm{d}[X \mid Z; Y \mid W] = \mathbf{H}[X' - Y' \mid Z', W'] - \tfrac{1}{2}\mathbf{H}[X' \mid Z'] - \tfrac{1}{2}\mathbf{H}[Y' \mid W']. \tag{5.15}$$

If one of the conditionings is trivial (for example, if W takes just one value) then we omit that variable and write, for instance, $\mathrm{d}[X \mid Z; Y]$.

Proposition 5.6. *Let $\pi : G \to H$ be a homomorphism. Then for any independent G-valued random variables X, Y, one has*

$$\mathrm{d}[X; Y] = \mathrm{d}[\pi(X); \pi(Y)] + \mathrm{d}[X \mid \pi(X); Y \mid \pi(Y)] + \mathbf{I}[X - Y : \pi(X), \pi(Y) \mid \pi(X) - \pi(Y)].$$

In particular, distance contracts under homomorphisms:

$$\mathrm{d}[\pi(X); \pi(Y)] \leqslant \mathrm{d}[X; Y] \tag{5.16}$$

Proof. Expanding the definition of distance, and using the conditional entropy chain rule

$$\mathbf{H}[X] = \mathbf{H}[\pi(X)] + \mathbf{H}[X \mid \pi(X)]$$

and

$$\mathbf{H}[Y] = \mathbf{H}[\pi(Y)] + \mathbf{H}[Y \mid \pi(Y)],$$

it suffices to establish the identity

$$\mathbf{H}[X - Y] = \mathbf{H}[\pi(X) - \pi(Y)] + \mathbf{H}[X - Y \mid \pi(X), \pi(Y)] + \mathbf{I}[X - Y : \pi(X), \pi(Y) \mid \pi(X) - \pi(Y)],$$

But from the chain rule again we have

$$\mathbf{H}[X - Y] = \mathbf{H}[\pi(X) - \pi(Y)] + \mathbf{H}[X - Y \mid \pi(X) - \pi(Y)],$$

which reduces matters to proving that

$$\mathbf{H}[X-Y \mid \pi(X)-\pi(Y)] - \mathbf{H}[X-Y \mid \pi(X),\pi(Y)] = \mathbf{I}[X-Y : \pi(X),\pi(Y) \mid \pi(X)-\pi(Y)].$$

This is an instance of $\mathbf{I}[A : B \mid C] = \mathbf{H}[A \mid C] - \mathbf{H}[A \mid B, C]$, taking $A = X-Y$, $B = \pi(X)-\pi(Y)$ and $C = (\pi(X),\pi(Y))$, and noting that $\mathbf{H}[A \mid B, C] = \mathbf{H}[A \mid C]$ since C determines B. □

5.3. **Entropic analogue of PFR.** In this subsection, we state an entropic analogue of Theorem 5.2 and deduce Theorem 5.2 from it.

Theorem 5.7 (Entropic PFR)**.** *There is an absolute constant C with the following property. Let X, Y be $\mathbf{F}_2^n$-valued random variables. Then there is some subgroup $H \leqslant \mathbf{F}_2^n$ such that $\mathrm{d}[X; U_H] \leqslant C\mathrm{d}[X;Y]$.*

Proof of Theorem 5.2, assuming Theorem 5.7. Let A, K be as in the statement of Theorem 5.2, that is to say $A \subset \mathbf{F}_2^n$ and $|A+A| \leqslant K|A|$. Let U_A be the uniform distribution on A, thus $\mathbf{H}[U_A] = \log |A|$. By (5.1) and the fact that $U_A + U_A$ is supported on $A+A$, $\mathbf{H}[U_A+U_A] \leqslant \log |A+A|$. The doubling condition $|A+A| \leqslant K|A|$ therefore gives

$$\mathrm{d}[U_A; U_A] \leqslant \log K. \tag{5.17}$$

By Entropic PFR (Theorem 5.7), we may thus find a subspace H of $\mathbf{F}_2^n$ such that

$$\mathrm{d}[U_A; U_H] \leqslant C \log K. \tag{5.18}$$

By Lemma 5.3 we conclude that $|\log |H| - \log |A|| \leqslant 2C \log K$, and so

$$K^{-2C}|A| \leqslant |H| \leqslant K^{2C}|A|. \tag{5.19}$$

From the definition of distance, and since $\mathbf{H}[U_A] = \log |A|$ and $\mathbf{H}[U_H] = \log |H|$, (5.18) is equivalent to

$$\mathbf{H}[U_A - U_H] \leqslant \log(|A|^{1/2}|H|^{1/2}) + C \log K.$$

By (5.2) we conclude the existence of a point $x_0 \in \mathbf{F}_2^n$ such that

$$p_{U_A-U_H}(x_0) \geqslant K^{-C}|A|^{-1/2}|H|^{-1/2},$$

or equivalently

$$|A \cap (H+x_0)| \geqslant K^{-C}|A|^{1/2}|H|^{1/2}.$$

Applying the Ruzsa covering lemma Lemma 3.2, we may thus cover A by at most

$$\frac{|A + (A \cap (H+x_0))|}{|A \cap (H+x_0)|} \leqslant \frac{K|A|}{K^{-C}|A|^{1/2}|H|^{1/2}} = K^{C+1}\frac{|A|^{1/2}}{|H|^{1/2}}$$

translates of $\big(A \cap (H+x_0)\big) - \big(A \cap (H+x_0)\big)$, which is contained in H. If $|H| \leqslant |A|$ then we are already done thanks to (5.19). If $|H| > |A|$ then we can cover H by at most $2|H|/|A|$ translates of a subspace H' of H with $|H'| \leqslant |A|$. We can thus cover A by at most

$$2K^{C+1}\frac{|H|^{1/2}}{|A|^{1/2}}$$

translates of H', and the claim again follows from (5.19). □

Remark. As usual the letter C may denote different absolute constants; the constant in Theorem 5.2 will be larger than the one in Theorem 5.7.

6. Proof of Polynomial Freiman-Ruzsa

In this section we prove the entropic form of Polynomial Freiman-Ruzsa (PFR), that is to say Theorem 5.7. As shown at the end of the last section, Theorem 5.2 then follows.

6.1. **An iterative strategy.** The general strategy will be to proceed by induction on $\mathrm{d}[X;Y]$. We will outline in more detail how such a strategy works in Proposition 6.2 below.

The 'base-case' for the induction is the case $\mathrm{d}[X;Y] = 0$.

Lemma 6.1. *Let X, Y be $\mathbf{F}_2^n$-valued random variables and suppose that $\mathrm{d}[X;Y] = 0$. Then there is a subgroup $H \leqslant \mathbf{F}_2^n$ such that $\mathrm{d}[X;U_H] = \mathrm{d}[Y;U_H] = 0$.*

Proof. By the triangle inequality Lemma 5.4 we have $\mathrm{d}[X;X] = 0$. Define H to be the set of all h such that $p_X(x) = p_X(x+h)$ for all $x \in X$. One can check immediately that H is a subgroup of G.

We claim that if $p_X(t), p_X(t') > 0$ then $t - t' \in H$. For this, we first observe that if X_1, X_2 are independent copies of X then $X_1 - X_2$ and X_1 are independent. Indeed, from the definition of distance $\mathbf{H}[X_1 - X_2] = \mathbf{H}[X_1] = \mathbf{H}[X_2]$, and therefore

$$\begin{aligned}\mathbf{I}[X_1 - X_2 : X_1] &= \mathbf{H}[X_1 - X_2] + \mathbf{H}[X_1] - \mathbf{H}[X_1 - X_2, X_1]\\ &= \mathbf{H}[X_1 - X_2] + \mathbf{H}[X_1] - \mathbf{H}[X_2, X_1] = 0\end{aligned}$$

since X_1, X_2 are independent. Here, in the middle step we used the fact that $(X_1 - X_2, X_1)$ and $X_1, X_2)$ bijectively determine each other and hence have the same entropy. Since mutual information vanishes only for independent variables, it follows that $X_1 - X_2$ and X_1 are independent.

Returning to the proof of the claim, let x be an arbitrary element in the range of X. Then

$$\begin{aligned}p_X(t)p_X(x) &= \mathbf{P}(X_1 = t, X_2 = x)\\ &= \mathbf{P}(X_1 - X_2 = t - x, X_1 = t) = \mathbf{P}(X_1 - X_2 = t - x)p_X(t),\end{aligned}$$

and so (since $p_X(t) > 0$)

$$p_X(x) = \mathbf{P}(X_1 - X_2 = t - x).$$

Similarly

$$p_X(x) = \mathbf{P}(X_1 - X_2 = t' - x).$$

Comparing the last two equations gives $p_X(x) = p_X(x + t - t')$, and so $t - t' \in H$ as claimed.

Now fix some x_0 with $p_X(x_0) > 0$. Then if x is any other point with $p_X(x) > 0$, since $x - x_0 \in H$ we have $p_X(x) = p_X(x_0 + (x - x_0)) = p_X(x_0)$. Conversely, if $x - x_0 \in H$ then $p_X(x) = p_X(x_0 + (x - x_0)) > 0$ and so $p_X(x) = p_X(x_0)$. It follows that X has the uniform distribution on $H + x_0$, or equivalently $X - x_0$ is uniformly distributed on H. Thus $0 = \mathrm{d}[X;X] = \mathrm{d}[X - x_0; X] = \mathrm{d}[U_H; X]$.

By the same argument applied to Y, there is a subgroup H' such that $\mathrm{d}[Y;U_{H'}] = 0$. By the triangle inequality we have $\mathrm{d}[U_H;U_{H'}] = 0$. From this it follows (Sheet 3, Q9) that $H = H'$. The proof of the lemma is complete. □

Now we state the key property that will make the induction work. The following proposition encodes the idea that there is a pair (X', Y') which at the same time (i) enjoys a 'distance decrement' meaning that $\mathrm{d}[X';Y']$ is appreciably less than $\mathrm{d}[X;Y]$ and (ii) is 'related to' (X,Y) in the sense that the distance from X to X' is somewhat bounded, and similarly for the distance from Y to Y'.

Proposition 6.2. *There is an absolute constant $\eta > 0$ such that the following holds. If X, Y are $\mathbf{F}_2^n$-valued random variables, then one can find $\mathbf{F}_2^n$-valued random variables X', Y' such that*

$$\mathrm{d}[X';Y'] + \eta(\mathrm{d}[X';X] + \mathrm{d}[Y';Y]) \leqslant (1-\eta)\mathrm{d}[X;Y]. \tag{6.1}$$

It is useful to note that the conclusion (6.1) implies

$$\mathrm{d}[X';Y'] \leqslant (1-\eta)\mathrm{d}[X;Y] \quad \text{and} \quad \mathrm{d}[X';X], \mathrm{d}[Y';Y] \leqslant \eta^{-1}\mathrm{d}[X;Y]. \tag{6.2}$$

The proof of Proposition 6.2 will be the main business of the section. Let us first see how it implies the entropic PFR, Theorem 5.7.

Proof of Theorem 5.7, assuming Proposition 6.2. Suppose, as in the statement of Theorem 5.7, that we have $\mathbf{F}_2^n$-valued random variables X, Y. By iterated application of Proposition 6.2, one can then find sequences of random variables X_n, Y_n with $X_0 = X$, $Y = Y_0$ and satisfying

$$\mathrm{d}[X_n; Y_n] \leqslant (1-\eta)^n \mathrm{d}[X; Y],$$

and

$$\mathrm{d}[X_{n+1}; X_n], \mathrm{d}[Y_{n+1}; Y_n] \leqslant \eta^{-1}(1-\eta)^n \mathrm{d}[X; Y].$$

In particular, from the triangle inequality and geometric series

$$\mathrm{d}[X_n; X], \mathrm{d}[Y_n; Y] \leqslant \eta^{-2} \mathrm{d}[X; Y].$$

The space of probability measures/random variables on $\mathbf{F}_2^n$ is compact, and so (passing to a subsequence of the X_n, Y_n) we may find limiting random variables X_∞, Y_∞ with $p_{X_\infty}(x) = \lim_{n\to\infty} p_{X_n}(x)$ for all x and $p_{Y_\infty(y)} = \lim_{n\to\infty} p_{Y_n}(y)$ (with the limits being along a subsequence). It is clear by inspection that the Ruzsa distance is continuous as a function on the space of probability measures, and so

$$\mathrm{d}[X_\infty; Y_\infty] = 0$$

and

$$\mathrm{d}[X_\infty; X], \mathrm{d}[Y_\infty; Y] \leqslant \eta^{-2} \mathrm{d}[X; Y]. \tag{6.3}$$

By Lemma 6.1, there is some subgroup H such that $\mathrm{d}[X_\infty; U_H] = \mathrm{d}[Y_\infty; U_H] = 0$. From (6.3) and the triangle inequality we then have $\mathrm{d}[X; U_H], \mathrm{d}[Y; U_H] \leqslant \eta^{-2} \mathrm{d}[X; Y]$, and the proof is complete. □

The remainder of the work to prove Theorem 5.7 is therefore in establishing Proposition 6.2.

6.2. **Using sums and fibres.** Let X, Y be $\mathbf{F}_2^n$-valued random variables, as in Proposition 6.2. We may assume without loss of generality that X, Y are independent. Let $\tilde{X}, \tilde{Y}$ be further independent copies. The idea now is to attempt to locate X', Y' satisfying (6.1) from among the following choices:

- sums $X'_* = X + \tilde{Y}$, $Y'_* = Y + \tilde{X}$;
- fibres $X'_t = (X \mid X + \tilde{Y} = t)$, $Y'_u = (Y \mid Y + \tilde{X} = u)$ for some t, u.

The crucial lemma is Lemma 6.4 below. This depends on the following corollary of the fibring identity, Proposition 5.6.

Corollary 6.3. *Let Z_1, Z_2, Z_3 and Z_4 be independent random variables taking values in $\mathbf{F}_2^n$. Then*

$$\begin{aligned} &\mathrm{d}[Z_1 + Z_3; Z_2 + Z_4] + \mathrm{d}[Z_1 \mid Z_1 + Z_3; Z_2 \mid Z_2 + Z_4] \\ &\quad + \mathbf{I}[Z_1 + Z_2 : Z_2 + Z_4 \mid Z_1 + Z_2 + Z_3 + Z_4] = \mathrm{d}[Z_1; Z_2] + \mathrm{d}[Z_3; Z_4]. \end{aligned}$$

Proof. We apply Proposition 5.6 with $G := \mathbf{F}_2^n \times \mathbf{F}_2^n$, $H := \mathbf{F}_2^n$, and π the homomorphism $\pi(x, y) := x + y$, and with the random variables $X := (Z_1, Z_3)$ and $Y := (Z_2, Z_4)$. (Note here that X, Y are dummy variables inside Proposition 5.6, and not the ones in Proposition 6.2). Then by independence we easily calculate

$$\mathrm{d}[X; Y] = \mathrm{d}[Z_1; Z_2] + \mathrm{d}[Z_3; Z_4]$$

while by definition

$$\mathrm{d}[\pi(X); \pi(Y)] = \mathrm{d}[Z_1 + Z_3; Z_2 + Z_4].$$

Furthermore,

$$\mathrm{d}[X \mid \pi(X); Y \mid \pi(Y)] = \mathrm{d}[Z_1 \mid Z_1 + Z_3; Z_2 \mid Z_2 + Z_4],$$

since $X = (Z_1, Z_3)$ and Z_1 are linked by an invertible affine transformation once $\pi(X) = Z_1 + Z_3$ is fixed, and similarly for Y and Z_2. Finally, we have

$$\begin{aligned}&\mathbf{I}[X+Y : (\pi(X), \pi(Y)) \mid \pi(X)+\pi(Y)]\\ &\quad = \mathbf{I}[(Z_1+Z_2, Z_3+Z_4) : (Z_1+Z_3, Z_2+Z_4) \mid Z_1+Z_2+Z_3+Z_4]\\ &\quad = \mathbf{I}[Z_1+Z_2 : Z_2+Z_4 \mid Z_1+Z_2+Z_3+Z_4]\end{aligned}$$

where in the last line we used the fact that $(Z_1+Z_2, Z_1+Z_2+Z_3+Z_4)$ uniquely determine Z_3+Z_4 and similarly $(Z_2+Z_4, Z_1+Z_2+Z_3+Z_4)$ uniquely determine Z_1+Z_3. □

For the remainder of the section, the following notation will be in place. X, Y are independent $\mathbf{F}_2^n$-valued random variables with $\mathrm{d}[X;Y] = k$, and $\tilde{X}, \tilde{Y}$ are further independent copies of them. Let $\eta > 0$ be a suitably small absolute constant as in the statement of Proposition 6.2. Recall the definition of *sums* and *fibres* from the start of Section 6.2.

Lemma 6.4. *Either the sums or the fibres (for some t, u) give a pair (X', Y') satisfying (6.1), or else we have the mutual information bound*

$$\mathbf{I}[X+Y : \tilde{X}+Y \mid X+Y+\tilde{X}+\tilde{Y}] = O(\eta k). \tag{6.4}$$

Proof. We apply Corollary 6.3 with the choice

$$(Z_1, Z_2, Z_3, Z_4) := (X, Y, \tilde{Y}, \tilde{X}).$$

It gives (recalling that we are in characteristic 2!) that

$$\mathrm{d}[X+\tilde{Y}; Y+\tilde{X}] + \mathrm{d}[X \mid X+\tilde{Y}; Y \mid Y+\tilde{X}] = 2k - I, \tag{6.5}$$

where I is the mutual information quantity on the LHS of (6.4). Recalling the definition of conditional distance, we can write this as

$$\mathrm{d}[X'_*; Y'_*] + \sum_{t,u} p_{X+\tilde{Y}}(t) p_{Y+\tilde{X}}(u) \mathrm{d}[X'_t; Y'_u] = 2k - I. \tag{6.6}$$

We claim that

$$\mathrm{d}[X'_*; X], \mathrm{d}[Y'_*; Y] = O(k) \tag{6.7}$$

and

$$\sum_{t,u} p_{X+\tilde{Y}}(t) p_{Y+\tilde{X}}(u) \mathrm{d}[X'_t; X] = O(k), \tag{6.8}$$

and similarly

$$\sum_{t,u} p_{X+\tilde{Y}}(t) p_{Y+\tilde{X}}(u) \mathrm{d}[Y'_u; Y] = O(k), \tag{6.9}$$

Assuming (6.7) to (6.9), it is easy to see using (6.6) that either the sums $(X', Y') = (X'_*, Y'_*)$ or the fibres $(X', Y') = (X'_t, Y'_u)$ (for some t, u) satisfy (6.1), or else $I = O(\eta k)$, which is the desired conclusion.

It remains to prove the three claims (6.7) to (6.9). These are exercises in 'entropic Ruzsa calculus'. We will prove the bound for $\mathrm{d}[X'_*; X]$ and (6.8) in detail, leaving the others (which are very similar) to the reader.

Proof of (6.7). We will in fact show that $\mathrm{d}[X'_*; X] = \mathrm{d}[X+\tilde{Y}; X] \leqslant 2k$. The key to this is the Madiman/Kaimanovich-Vershik inequality Lemma 5.5.

Writing X_1, X_2 for independent copies of X,

$$\mathrm{d}[X; X+\tilde{Y}] = \mathbf{H}[X_1+Y+X_2] - \tfrac{1}{2}\mathbf{H}[X_1+Y] - \tfrac{1}{2}\mathbf{H}[X].$$

By Lemma 5.5 we thus have

$$\begin{aligned} \mathrm{d}[X; X+\tilde{Y}] &\leqslant \mathbf{H}[X_1+Y]+\mathbf{H}[X_2+Y]-\mathbf{H}[Y]-\tfrac{1}{2}\mathbf{H}[X_1+Y]-\tfrac{1}{2}\mathbf{H}[X] \\ &= \tfrac{3}{2}\mathbf{H}[X+Y]-\mathbf{H}[Y]-\tfrac{1}{2}\mathbf{H}[X] \\ &= \tfrac{3}{2}\mathrm{d}[X;Y]+\tfrac{1}{4}(\mathbf{H}[X]-\mathbf{H}[Y]). \end{aligned}$$

By Lemma 5.3 and the definition of distance, this is bounded by $2k$, as desired.

Proof of (6.8). Recalling the definition of conditional distance, the left-hand size is $\mathrm{d}[X \mid X+\tilde{Y}; X]$, and we will show that this is at most $3k$. We will use the following inequality, which is Sheet 4, Q1.

$$\mathrm{d}[A \mid Z; B \mid W] \leqslant \mathrm{d}[A;B]+\tfrac{1}{2}(\mathbf{I}[A:Z]+\mathbf{I}[B:W]) \tag{6.10}$$

Applying this with W trivial gives

$$\mathrm{d}[A \mid Z; B] \leqslant \mathrm{d}[A;B]+\tfrac{1}{2}\mathbf{I}[A:Z] \tag{6.11}$$

Taking $A=B=X$ and $Z=X+Y$, we obtain

$$\mathrm{d}[X \mid X+\tilde{Y}; X] \leqslant \mathrm{d}[X;X]+\tfrac{1}{2}\mathbf{I}[X, X+Y]. \tag{6.12}$$

By the triangle inequality, $\mathrm{d}[X;X] \leqslant 2k$. Finally,

$$\begin{aligned} \mathbf{I}[X, X+Y] &= \mathbf{H}[X]+\mathbf{H}[X+Y]-\mathbf{H}[X, X+Y] \\ &= \mathbf{H}[X]+\mathbf{H}[X+Y]-\mathbf{H}[X,Y] \\ &= \mathbf{H}[X+Y]-\mathbf{H}[X] \\ &= \mathrm{d}[X;Y]+\tfrac{1}{2}(\mathbf{H}[Y]-\mathbf{H}[X]). \end{aligned}$$

By another application of Lemma 5.3, this is $\leqslant 2k$. Combining with (6.12), we obtain the desired result. □

We in fact need two slight variants of Lemma 6.4, proven using slight variants of the sums and fibres as described above. The following is the only result we will need going forwards.

Lemma 6.5. *Either there is a pair (X', Y') satisfying* (6.1)*, or else we have, writing $S := X+Y+\tilde{X}+\tilde{Y}$:*

$$\mathbf{I}[X+Y : \tilde{X}+Y \mid S] = O(\eta k). \tag{6.13}$$

$$\mathbf{I}[X+Y : X+\tilde{X} \mid S] = O(\eta k). \tag{6.14}$$

$$\mathbf{I}[\tilde{X}+Y : X+\tilde{X} \mid S] = O(\eta k), \tag{6.15}$$

Proof. We have already proven (6.13) in Lemma 6.4 above. (6.14) and (6.15) are equivalent after swapping tildes. Thus it remains to establish (6.14). This is done via a mild variant of Lemma 6.4, in which we instead apply Corollary 6.3 with the choice $(Z_1, Z_2, Z_3, Z_4) = (Y, X, \tilde{Y}, \tilde{X})$. The rest of the proof is a minor variant of that of Lemma 6.4, with differences in the details of the proofs of the analogues (6.7) to (6.9). One or two of these details are on Sheet 4; the rest are left to the interested reader. □

6.3. **Finishing the argument: entropic BSG.** The aim in this section is to complete the proof of PFR by locating an appropriate pair (X', Y') of random variables satisfying (6.1), assuming that the three conditions (6.13) to (6.15) all hold. Write

$$U := X+Y, \qquad V := \tilde{X}+Y, \qquad W := X+\tilde{X}$$

and recall that

$$S := X+Y+\tilde{X}+\tilde{Y}.$$

Then the aforementioned three conditions may be written (respectively) as

$$\mathbf{I}[U : V \mid S], \mathbf{I}[U : W \mid S], \mathbf{I}[V : W \mid S] = O(\eta k). \tag{6.16}$$

At this point we make the only critical use of characteristic 2 in the proof, noting the crucial relation

$$U + V + W = 0.$$

(Whilst we did make an earlier appeal to characteristic 2 in order to replaced some minus signs by plusses, this was not critical to the argument.)

Let us ignore for a moment the conditioning upon S, and let us also suppose that (6.16) is replaced by the stronger (but similar) statement that U, V, W are independent. Then we have (remembering we are in characteristic 2)

$$\mathrm{d}[U; V] = \mathbf{H}[U + V] - \tfrac{1}{2}\mathbf{H}[U] - \tfrac{1}{2}\mathbf{H}[V] = \mathbf{H}[W] - \tfrac{1}{2}\mathbf{H}[U] - \tfrac{1}{2}\mathbf{H}[V]$$

and similar relations cyclically. Adding these relations gives

$$\mathrm{d}[U; V] + \mathrm{d}[V; W] + \mathrm{d}[W; U] = 0,$$

so all the three distance are zero, which means U, V, W are uniform on cosets of a subgroup by Lemma 6.1. This would (in this idealised situation) put us in a very strong position to conclude the proof of PFR.

To make this rigorous we need to do three things:

- Take account of the conditioning by S;
- Deal with the fact that (6.16) does *not* give true independence of the variables U, V, W (conditioned on S);
- Show that appropriate variables are reasonably close in distance to X, Y.

To handle the conditioning we use the inequality

$$\max(\mathbf{H}[U \mid S], \mathbf{H}[V \mid S]) - \mathbf{I}[U : V \mid S] \leqslant \mathbf{H}[U + V \mid S].$$

The unconditioned version of this was Sheet 3, Q3; the conditioned version follows from it by applying Sheet 3, Q3 to $(U \mid S = s)$ and $(V \mid S = s)$ and summing over s, weighted by $p_S(s)$. Since $U + V = W$, it follows from this and (6.16) that

$$\mathbf{H}[U \mid S], \mathbf{H}[V \mid S] \leqslant \mathbf{H}[W \mid S] + O(\eta k).$$

The same holds for cyclic permutations of the variables, so all three of $\mathbf{H}[U \mid S], \mathbf{H}[V \mid S], \mathbf{H}[W \mid S]$ differ by $O(\eta k)$ at most. In particular, since $U + V = W$ we have

$$\mathbf{H}[U + V \mid S] - \tfrac{1}{2}\mathbf{H}[U \mid S] - \tfrac{1}{2}\mathbf{H}[V \mid S] \leqslant O(\eta k). \tag{6.17}$$

Writing $U_s := (U \mid S = s)$ (and similarly) for brevity, this expands as

$$\sum_s p_S(s)\Big(\mathbf{H}[U_s + V_s] - \tfrac{1}{2}\mathbf{H}[U_s] - \tfrac{1}{2}\mathbf{H}[V_s]\Big) = O(\eta k). \tag{6.18}$$

Note carefully that we do *not* write $\mathrm{d}[U_s; V_s]$ for the bracketed expression, since U_s, V_s are not independent (though the bound (6.16) implies that they are nearly so in an average sense). The bound $\mathbf{I}[U : V|S] = O(\eta k)$ expands as

$$\sum_s p_S(s)\mathbf{I}[U_s : V_s] = O(\eta k). \tag{6.19}$$

To deal with the lack of independence of U_s, V_s we invoke a tool called the Entropic Balog-Szemerédi-Gowers theorem. We give the statement only, outsourcing the proof (together with the explanation of the name) to Appendix B.

Lemma 6.6 (Entropic BSG)**.** *Let A, B be (not necessarily independent) G-valued random variables. Then*

$$\sum_z p_Z(z)\mathrm{d}[A'_z; B'_z] \leqslant 3\mathbf{I}[A : B] + 2\big(\mathbf{H}[A+B] - \tfrac{1}{2}\mathbf{H}[A] - \tfrac{1}{2}\mathbf{H}[B]\big),$$

where $A'_z := (A \mid A+B = z)$, $B'_z := (B \mid A+B = z)$.

The point here is that we can take an assumption on $\mathbf{H}[A+B] - \frac{1}{2}\mathbf{H}[A] - \frac{1}{2}\mathbf{H}[B]$ being small and, provided A, B are almost independent, convert it to random variables A', B' at small distance.

We may apply Lemma 6.6 in our situation with $A = U_s$, $B = V_s$, so $A + B = W_s$. Write $X'_{s,w} := (U_s \mid W_s = w)$ and $Y'_{s,w} := (V_s \mid W_s = w)$. By Lemma 6.6, (6.18), and (6.19) we have (weighting by $p_S(s)$ and summing),

$$\sum_{s,w} p_S(s)p_{W_s}(w)\mathrm{d}[X'_{s,w}; Y'_{s,w}] = O(\eta k). \tag{6.20}$$

If η is small, we have produced some variables at distance significantly smaller than $k = \mathrm{d}[X; Y]$. To conclude the argument, we need to show that they are (on average) not too distant from X and Y. Indeed, suppose we can show

$$\sum_{s,w} p_S(s)p_{W_s}(w)\mathrm{d}[X'_{s,w}; X] = O(k) \tag{6.21}$$

and

$$\sum_{s,w} p_S(s)p_{W_s}(w)\mathrm{d}[Y'_{s,w}; Y] = O(k). \tag{6.22}$$

Then

$$\sum_{s,w} p_S(s)p_{W_s}(w)\Big(\mathrm{d}[X'_{s,w}; Y'_{s,w}] + \eta\mathrm{d}[X'_{s,w}; X] + \eta\mathrm{d}[Y'_{s,w}; Y]\Big) = O(\eta k),$$

so (provided η is a sufficiently small absolute constant) the LHS is $\leqslant (1-\eta)k$. In particular, there is some choice of $X' = X'_{s,w}$ and $Y' = Y'_{s,w}$ such that (6.1) holds, thereby concluding the proof of Proposition 6.2.

It remains to establish (6.21) and (6.22). The proofs are similar so we handle only (6.21). By (6.11) (which follows immediately from (6.10), which was an exercise on Sheet 4) we have

$$\sum_w p_{W_s}(w)\mathrm{d}[X; X'_{s,w}] = \mathrm{d}[X; U_s \mid W_s] \leqslant \mathrm{d}[X; U_s] + \tfrac{1}{2}\mathbf{I}[U_s : W_s].$$

Summing over s we obtain

$$\sum_s p_S(s) \sum_w p_{W_s}(w)\mathrm{d}[X'_{s,w}; X] \leqslant \mathrm{d}[X; U \mid S] + \tfrac{1}{2}\mathbf{I}[U : W \mid S].$$

By (6.16), it remains only to show that

$$\mathrm{d}[X; U \mid S] = O(k).$$

This is an exercise in entropic Ruzsa calculus and is covered in Sheet 4, Q2.

7. The weak PFR in the integers

In this section we use Theorem 5.7 to prove a result about the integers called the weak PFR.

Theorem 7.1 (Weak PFR in $\mathbf{Z}$)**.** *There is an absolute constant C_2 such that the following is true. Suppose that $A \subseteq \mathbf{Z}^D$ is a finite set with $\sigma[A] \leqslant K$. Then there is $A' \subseteq A$ with $|A'| \geqslant K^{-C_2}|A|$ and $\dim A' \leqslant C_2 \log K$.*

Remark. The term 'weak' comes from the fact that we only control the dimension of A, and not anything more precise (for instance putting A inside a progression).

7.1. **Projections modulo 2.** A key idea is that a set $A \subseteq \mathbf{Z}^D$ with small doubling must look rather singular under the projection map $\phi : \mathbf{Z}^D \to \mathbf{F}_2^D$. In Lemma 7.3 below, we give an entropic formulation of this principle (followed by an attempt to explain why we expect it to hold). We isolate the following lemma from the proof.

Lemma 7.2. *Let G be torsion-free, and let X, Y be G-valued random variables. Then* $\mathrm{d}[X; 2Y] \leqslant 5\mathrm{d}[X; Y]$.

Proof. The proof is a little tricky. We may assume X, Y are independent. At this point we use the slightly stronger version of the triangle inequality, (5.13), taking $A = X - Y$, $B = C = Y$ in that inequality. This gives

$$\begin{aligned}\mathbf{H}[X - 2Y] = \mathbf{H}[(X - Y) - Y] &\leqslant \mathrm{d}[Y; Y] + \mathrm{d}[X - Y; Y] + \frac{1}{2}\mathbf{H}[X - Y] + \frac{1}{2}\mathbf{H}[Y] \\ &\leqslant 2\mathrm{d}[X; Y] + \mathrm{d}[X - Y; Y] + \frac{1}{2}\mathbf{H}[X - Y] + \frac{1}{2}\mathbf{H}[Y], \qquad (7.1)\end{aligned}$$

where in the final step we used the triangle inequality again.

Let Y_1, Y_2 be independent copies of Y (which are also independent of X). Then we have

$$\mathrm{d}[X - Y; Y] = \mathbf{H}[X - Y_1 - Y_2] - \frac{1}{2}\mathbf{H}[X - Y] - \frac{1}{2}\mathbf{H}[Y]. \qquad (7.2)$$

Writing $A := Y_1$, $B := Y_2$ and $C := X - Y_1 - Y_2$, we have

$$\mathbf{H}[A, B, C] = \mathbf{H}[X, Y_1, Y_2] = \mathbf{H}[X] + 2\mathbf{H}[Y],$$

and

$$\begin{aligned}\mathbf{H}[A, C] = \mathbf{H}[A, C + A] = \mathbf{H}[Y_1, X - Y_2] = \mathbf{H}[Y] + \mathbf{H}[X - Y_2] = \mathbf{H}[Y] + \mathbf{H}[X - Y], \\ \mathbf{H}[B, C] = \mathbf{H}[B, C + B] = \mathbf{H}[Y_2, X - Y_1] = \mathbf{H}[Y] + \mathbf{H}[X - Y_1] = \mathbf{H}[Y] + \mathbf{H}[X - Y]\end{aligned}$$

so applying the submodularity inequality (5.8) gives

$$\mathbf{H}[X - Y_1 - Y_2] \leqslant 2\mathbf{H}[X - Y] - \mathbf{H}[X].$$

Combining this with (7.2) gives

$$\mathrm{d}[X - Y; Y] \leqslant \frac{3}{2}\mathbf{H}[X - Y] - \mathbf{H}[X] - \frac{1}{2}\mathbf{H}[Y].$$

From (7.1) it follows that

$$\mathbf{H}[X - 2Y] \leqslant 2\mathrm{d}[X; Y] + 2\mathbf{H}[X - Y] - \mathbf{H}[X] = 4\mathrm{d}[X; Y] + \mathbf{H}[Y].$$

Finally, we obtain

$$\mathrm{d}[X; 2Y] \leqslant 4\mathrm{d}[X; Y] + \frac{1}{2}(\mathbf{H}[Y] - \mathbf{H}[X]) \leqslant 5\mathrm{d}[X; Y]$$

where we used Lemma 5.3 in the last step. □

Lemma 7.3. *Let X, Y be $\mathbf{Z}^D$-valued random variables for some $D \geqslant 0$. Denote by $\phi : \mathbf{Z}^D \to \mathbf{F}_2^D$ the natural homomorphism. Then*

$$\mathbf{H}[\phi(X)], \mathbf{H}[\phi(Y)] \leqslant 10\mathrm{d}[X; Y].$$

Proof. By the contraction of distance under homomorphisms (5.16) and Lemma 7.2,

$$\mathrm{d}[\phi(X); \phi(2Y)] \leqslant \mathrm{d}[X; 2Y] \leqslant 5\mathrm{d}[X; Y]. \qquad (7.3)$$

However, $\phi(2Y)$ is identically zero and so

$$\mathrm{d}[\phi(X); \phi(2Y)] = \mathrm{d}[\phi(X); 0] = \frac{1}{2}\mathbf{H}[\phi(X)].$$

Combining this with (7.3) gives the stated bound for $\mathbf{H}[\phi(X)]$. The bound for $\mathbf{H}[\phi(Y)]$ follows in the same way. □

Remark. It is perhaps worth remarking on the meaning and proof of this statement. Supposing that $A \subset \mathbf{Z}^D$ is a set with small (combinatorial) doubling K, it follows that the dilate $2 \cdot A$, which is contained in $A + A$, is commensurate (up to polynomial factors in K) with A. Projecting mod 2, one therefore expects the projection $\pi(A)$ to be commensurate with the projection $\pi(2 \cdot A) = \{0\}$. In the entropy setting, Lemma 7.2 acts as a replacement for the trivial observation that $2 \cdot A$ is contained in $A + A$.

7.2. **Projections and iterated PFR.** The main result of this subsection is the slightly technical Lemma 7.5 below, which is obtained by iterated application of PFR. It states that if X, Y are $\mathbf{F}_2^n$-valued random variables with small distance, then in some sense we may 'capture most of the entropy' of X, Y by projecting out a reasonably small subspace H.

During the proof we will need a basic fact about the behaviour of entropy under group homomorphism, which we detail now.

Lemma 7.4. *Let X be a G-valued random variable, and let H be a finite subgroup of G. Denote by $\pi : G \to G/H$ the quotient map. Let U_H be a uniform random variable on H. Then $\mathbf{H}[\pi(X)] \leqslant 2\mathrm{d}[X; U_H]$.*

Proof. We first observe that

$$\mathbf{H}[X - U_H] = \mathbf{H}[X + U_H] = \mathbf{H}[\pi(X)] + \mathbf{H}[U_H] = \mathbf{H}[\pi(X)] + \log |H|. \tag{7.4}$$

The first equality follows from the fact that $H = -H$, and the third is immediate. Only the middle equality needs further explanation. For this, pick a 'section' of G over π, that is to say a choice of elements $(g_t)_{t \in G/H}$ with $\pi(g_t) = t$ for all t. Now define a random variable Y by sampling t from $\pi(X)$ and setting $Y = g_t$. By (5.3) we have $\mathbf{H}[Y] = \mathbf{H}[\pi(X)]$. Also, if U_H is a uniform random on H, independent of Y, we see that $Y + U_H$ and $X + U_H$ have the same distribution. Moreover, the natural map $(Y, U_H) \mapsto Y + U_H$ is injective. Therefore

$$\mathbf{H}[X + U_H] = \mathbf{H}[Y + U_H] = \mathbf{H}[Y, U_H] = \mathbf{H}[Y] + \mathbf{H}[U_H] = \mathbf{H}[\pi(X)] + \mathbf{H}[U_H],$$

as stated.

It follows from (7.4) that

$$\mathrm{d}[X; U_H] = \mathbf{H}[\pi(X)] + \frac{1}{2}(\log |H| - \mathbf{H}[X]), \tag{7.5}$$

and the lemma then follows using Lemma 5.3. □

Now we state the main result of this subsection. Let C be the implied constant in PFR (Theorem 5.7).

Lemma 7.5. *Suppose that X and Y are $\mathbf{F}_2^D$-valued random variables. Then there is a subgroup $H \leqslant \mathbf{F}_2^D$ such that, denoting by $\psi \colon \mathbf{F}_2^D \to \mathbf{F}_2^D/H$ the natural projection, we have*

$$\log |H| \leqslant 2(\mathbf{H}[X] + \mathbf{H}[Y]) \tag{7.6}$$

and

$$\mathbf{H}[\psi(X)] + \mathbf{H}[\psi(Y)] \leqslant 8C\mathrm{d}[\psi(X); \psi(Y)]. \tag{7.7}$$

We isolate the following (sub-) lemma from the proof. This results from a direct application of PFR, together with understanding the behaviour of entropy under projections.

Lemma 7.6. *Let $n \in \mathbf{N}$. Let X, Y be $\mathbf{F}_2^n$-valued random variables. Set $k := \mathrm{d}[X;Y]$, and suppose that*

$$\mathbf{H}[X] + \mathbf{H}[Y] > 8Ck. \tag{7.8}$$

Then there is a nontrivial subgroup $H \leqslant \mathbf{F}_2^n$ such that

$$\log |H| \leqslant \mathbf{H}[X] + \mathbf{H}[Y] \tag{7.9}$$

and (writing $\psi\colon \mathbf{F}_2^n \to \mathbf{F}_2^n/H$ for the natural projection)

$$\mathbf{H}[\psi(X)] + \mathbf{H}[\psi(Y)] \leqslant \frac{1}{2}\big(\mathbf{H}[X] + \mathbf{H}[Y]\big). \tag{7.10}$$

Proof. Set $k := \mathrm{d}[X;Y]$. Applying PFR (Theorem 5.7), we obtain a subgroup H such that $\mathrm{d}[X;U_H], \mathrm{d}[Y;U_H] \leqslant Ck$. By Lemma 7.4 and (7.8), it follows that

$$\mathbf{H}[\psi(X)] + \mathbf{H}[\psi(Y)] \leqslant 4Ck < \frac{1}{2}(\mathbf{H}[X] + \mathbf{H}[Y]),$$

which is (7.10). To prove (7.9), first note that an application of Lemma 5.3 yields

$$\log |H| - \mathbf{H}[X] \leqslant 2\mathrm{d}[X;U_H] \leqslant 2Ck,$$

and similarly for Y. Therefore using (7.8) we have

$$\log |H| \leqslant \frac{1}{2}(\mathbf{H}[X] + \mathbf{H}[Y]) + 2Ck < \mathbf{H}[X] + \mathbf{H}[Y],$$

which gives the required bound (7.9).

Finally, we need to prove that H is not trivial. If H were trivial we would have $\psi(X) = X$, $\psi(Y) = Y$ and so (7.10) would imply $\mathbf{H}[X] + \mathbf{H}[Y] = 0$, which is contrary to the assumption (7.8). □

Proof of Lemma 7.5. We iteratively define a sequence $\{0\} = H_0 < H_1 < \cdots$ of subgroups of $\mathbf{F}_2^D$. Denote by $\psi_i : \mathbf{F}_2^D \to \mathbf{F}_2^D/H_i$ the ith associated projection operator, and set $k_i := \mathrm{d}[\psi_i(X);\psi_i(Y)]$. We stop the iteration at the ith stage if we have

$$\mathbf{H}[\psi_i(X)] + \mathbf{H}[\psi_i(Y)] \leqslant 8Ck_i. \tag{7.11}$$

Otherwise, we apply Lemma 7.6 to $\psi_i(X), \psi_i(Y)$, obtaining a nontrivial subgroup $H_{i+1}/H_i \leqslant \mathbf{F}_2^D/H_i$ such that

$$\log \frac{|H_{i+1}|}{|H_i|} \leqslant \mathbf{H}[\psi_i(X)] + \mathbf{H}[\psi_i(Y)] \tag{7.12}$$

and

$$\mathbf{H}[\psi_{i+1}(X)] + \mathbf{H}[\psi_{i+1}(Y)] \leqslant \frac{1}{2}\big(\mathbf{H}[\psi_i(X)] + \mathbf{H}[\psi_i(Y)]\big). \tag{7.13}$$

Clearly from iterated application of (7.13) we obtain

$$\mathbf{H}[\psi_i(X)] + \mathbf{H}[\psi_i(Y)] \leqslant 2^{-i}(\mathbf{H}[X] + \mathbf{H}[Y]).$$

Then, from a telescoping application of (7.12) we get

$$\log |H_i| \leqslant 2(\mathbf{H}[X] + \mathbf{H}[Y]). \tag{7.14}$$

Since the groups H_i form a strictly increasing sequence, the iteration does terminate at some time i. At this time we have both (7.11) and (7.14) and so, setting $\psi = \psi_i$, the proof of Lemma 7.5 is concluded. □

7.3. **Proof of weak PFR in the integers.** Now we turn our attention to the weak PFR itself, Theorem 7.1. It is a consequence of the following bipartite statement. The bipartite statement is amenable to a proof by induction, as we shall see. Passing to a bipartite statement in something like this manner is often helpful, and is an example of a situation where it is easier to prove a stronger statement by induction.

Theorem 7.7. *There is an absolute constant C_1 such that the following is true. Let $D \in \mathbf{N}$, and suppose $A, B \subseteq \mathbf{Z}^D$ are finite non-empty sets, and set $k := \mathrm{d}[U_A; U_B]$. Then there exist nonempty $A' \subseteq A$, $B' \subseteq B$ with*

$$\log \frac{|A|}{|A'|} + \log \frac{|B|}{|B'|} \leqslant C_1 k$$

and such that $\dim A', \dim B' \leqslant C_1 k$.

Before giving the proof, let us see how Theorem 7.1 follows from it. Suppose that $A \subset \mathbf{Z}^D$ is a set with $\sigma[A] \leqslant K$. By the Ruzsa triangle inequality for sets (Lemma 3.1) with $V = W = A$ and $U = -A$ we have $|A - A| \leqslant K^2|A|$. By Sheet 3, Q9 (see also the analysis leading up to (5.17), with suitable minus signs) $\mathrm{d}[U_A; U_A] \leqslant 2\log K$. Apply Theorem 7.7 with $A = B$ and $k = 2\log K$; we then obtain $A', A'' \subset A$ with

$$\log \frac{|A|}{|A'|} + \log \frac{|A|}{|A''|} \leqslant 2C_1 \log K$$

such that $\dim A', \dim A'' \leqslant 2C_1 \log K$. In particular, $|A'| \geqslant K^{-2C_1}|A|$ and the conclusion of Theorem 7.1 follows (with $C_2 = 2C_1$).

Proof of Theorem 7.7. Take $C_1 := \max(\frac{40}{\log 2}, 8C)$. We will proceed by induction on $|A| + |B|$. We may also assume that A, B do not sit inside cosets of a proper subgroup of $\mathbf{Z}^D$, else we may replace $\mathbf{Z}^D$ by that subgroup (which is still isomorphic to some $\mathbf{Z}^{D'}$).

Let $\phi : \mathbf{Z}^D \to \mathbf{F}_2^D$ be the natural homomorphism. By Lemma 7.3 we have

$$\mathbf{H}[\phi(U_A)], \mathbf{H}[\phi(U_B)] \leqslant 10k. \tag{7.15}$$

Applying Lemma 7.5 to $\phi(U_A), \phi(U_B)$, we find a subgroup $H \leqslant \mathbf{F}_2^D$ and associated projection $\psi : \mathbf{F}_2^D \to \mathbf{F}_2^D/H$ such that, denoting by $\tilde{\phi} = \psi \circ \phi : \mathbf{Z}^D \to \mathbf{F}_2^D/H$ the natural (composite) projection, we have

$$\log |H| \leqslant 2(\mathbf{H}[\phi(U_A)] + \mathbf{H}[\phi(U_B))] \leqslant 40k \tag{7.16}$$

and

$$M \leqslant 8C\tilde{k}, \tag{7.17}$$

where

$$M := \mathbf{H}[\tilde{\phi}(U_A)] + \mathbf{H}[\tilde{\phi}(U_B)] \quad \text{and} \quad \tilde{k} := \mathrm{d}[\tilde{\phi}(U_A); \tilde{\phi}(U_B)]. \tag{7.18}$$

Now if H is all of $\mathbf{F}_2^D$ then it follows from (7.16) that $D \leqslant \frac{40}{\log 2}k$, and so Theorem 7.7 is true simply by taking $A' = A$, $B' = B$, by the choice of C_1.

Suppose, then, that H is not all of $\mathbf{F}_2^D$. For $x, y \in \mathbf{F}_2^D/H$, denote by $A_x := A \cap \tilde{\phi}^{-1}(x)$ and $B_y := B \cap \tilde{\phi}^{-1}(y)$ the fibres of A, B above x, y respectively. Since we are assuming that A, B do not sit inside cosets of a proper subgroup of $\mathbf{Z}^D$, we may assume that at least one of $\tilde{\phi}(A), \tilde{\phi}(B)$ is not a singleton, and so

$$|A_x| + |B_y| < |A| + |B| \tag{7.19}$$

and $M > 0$, whereby $\tilde{k} > 0$ by (7.17).

Now we apply the fibring identity, Proposition 5.6. Ignoring the $\mathbf{I}[\,]$-term, which is non-negative, this implies that

$$\mathrm{d}[U_A \mid \phi(U_A); U_B \mid \phi(U_B)] \leqslant k - \tilde{k}. \tag{7.20}$$

It follows from this and (7.17) that

$$\sum_{x,y}\alpha_x\beta_y\log(1/\alpha_x\beta_y)+8C\sum_{x,y}\alpha_x\beta_y\mathrm{d}[U_{A_x};U_{B_y}]=M+8C\mathrm{d}[U_A\mid\phi(U_A);U_B\mid\phi(U_B)]$$
$$\leqslant 8C\tilde{k}+8C(k-\tilde{k})=8Ck.$$

Since $\sum_{x,y}\alpha_x\beta_y=1$, it follows that there are $x,y\in\mathbf{F}_2^D/H$ such that

$$\log(1/\alpha_x\beta_y)=\log\frac{|A|}{|A_x|}+\log\frac{|B|}{|B_y|}\leqslant 8C\big(k-\mathrm{d}[U_{A_x};U_{B_y}]\big). \tag{7.21}$$

Fix such a choice of x,y and set $k'=\mathrm{d}[U_{A_x};U_{B_y}]$. It follows from (7.21) that $k'\leqslant k$. By induction (and (7.19)) we may find $A'\subseteq A_x$ and $B'\subseteq B_y$ such that $\dim A',\dim B'\leqslant C_1k'\leqslant C_1k$ and

$$\log\frac{|A_x|}{|A'|}+\log\frac{|B_y|}{|B'|}\leqslant C_1k'.$$

Adding this to (7.21) yields

$$\log\frac{|A|}{|A'|}+\log\frac{|B|}{|B'|}\leqslant C_1k'+8C(k-k')\leqslant C_1k \tag{7.22}$$

since $C_1\geqslant 8C$. This closes the induction and the proof is complete. □

8. Combinatorial geometry and sum-product

Let A be a set of n integers. We have already discussed the sumset $A+A$ at some length. We may also introduce the product set $A\cdot A:=\{aa':a,a'\in A\}$. A famous conjecture of Erdős and Szemerédi is that

$$|A+A|+|A\cdot A|\geqslant n^{2-o(1)}.$$

This is far from being proven, but the final result of this chapter is the non-trivial result

$$|A+A|+|A\cdot A|\gg n^{5/4},$$

which is due to Elekes. The main input in establishing this is the so-called Szemerédi-Trotter theorem, a result in combinatorial geometry of substantial independent interest.

Theorem 8.1 (Szemerédi-Trotter)**.** *Let $r\geqslant 2$. Let L be a set of m lines. Then the number of points which lie on at least r lines in L is $O(\frac{m}{r}+\frac{m^2}{r^3})$.*

There are various slightly different ways to state this theorem, a matter we discuss on the example sheets. The proof we shall give of this uses a lemma about crossing numbers which is also of independent interest.

8.1. Crossing number inequality.

This section assumes that you are familiar with the basic language of graph theory; if not, it should be easy to read up on the relevant definitions.

Definition 8.2. A *drawing* of a graph G is a representation of G in the plane $\mathbf{R}^2$ where the vertices of G are points and the edges are "nice" simple curves between pairs of vertices, not passing through any other vertex of the graph. A *crossing* is an intersection of two edge-curves, other than at a vertex. The *crossing number* cross(G) of a graph G is the least number of crossings in any drawing of G in the plane. A graph is said to be *planar* if cross$(G)=0$.

Remark. We will not bother to set up what "nice" means rigorously, and it does not really matter; for example, we could take the curves to be polygonal. Note also that crossings are counted as

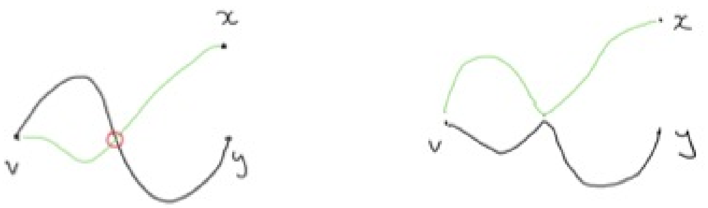

FIGURE 1. Removing a crossing

pairs of edge-curves which intersect, not as the actual points of intersection. Thus, for example, three edge-curves all intersecting at the same point counts as three crossings.

We begin by recalling *Euler's formula*. If G is a connected planar graph then

$$V - E + F = 2, \tag{8.1}$$

where V, E, F denote the numbers of vertices, edges and faces respectively. Now if $V \geqslant 3$ then every face has at least three edges, and no edge belongs to more than two faces. Therefore, double counting edges,

$$3F \leqslant 2E.$$

Substituting into Euler's formula (8.1) gives $E - 3V \leqslant -6$. Considering the cases where $V = 1$ or 2, one sees that certainly

$$E \leqslant 3V \tag{8.2}$$

in all cases. By splitting into connected components, we see that (8.2) holds for all planar graphs, connected or not.

Remark. Formalising the details here (even defining exactly what is meant by a face, especially in degenerate cases such as when G is a tree) is slightly subtle and not the domain of this course. For a much fuller discussion, see the graph theory course.

If we have a graph G then consider a drawing of G with $\text{cross}(G)$ crossings. For each such crossing, remove one of the edges in it. Continuing in this fashion gives a planar graph G' with the same vertex set as G and with $E' \geqslant E - \text{cross}(G)$ edges. It follows from (8.2) that $E' \leqslant 3V$ and so

$$\text{cross}(G) \geqslant E - 3V. \tag{8.3}$$

It turns out that by a random sampling trick we can bootstrap this to the following inequality, which is much stronger when E is relatively large in terms of V.

Proposition 8.3. *Suppose that $E \geqslant 4V$. Then* $\text{cross}(G) \geqslant \frac{E^3}{64V^2}$.

Proof. Take a drawing of G with the minimal number $\text{cross}(G)$ of crossings. Then all crossings involve four distinct vertices: if there is some crossing involving edges vx, vy then there is an easy procedure to reduce the number of crossings, best described by a picture (see the figure below).

Let p, $0 \leqslant p \leqslant 1$, be a parameter to be specified later. Consider a random subgraph $\mathbf{G}'$ of G, formed by picking a random set $\mathbf{S}$ of vertices by selecting each v in the vertex set of G to lie in $\mathbf{S}$

independently at random with probability p, and then taking $\mathbf{G}'$ to be the subgraph of G induced by $\mathbf{S}$: that is, include all the edges in G between two vertices in $\mathbf{S}$. Let the number of vertices and edges in $\mathbf{G}'$ be $\mathbf{V}'$, $\mathbf{E}'$ respectively; these are of course random variables. Also, let $\widetilde{\mathrm{cross}}(\mathbf{G}')$ be the number of crossings in $\mathbf{G}'$ in the drawing we have, that is to say in the drawing induced from that on G. Note that $\widetilde{\mathrm{cross}}(\mathbf{G}') \geqslant \mathrm{cross}(\mathbf{G}')$, but we do not necessarily have equality since there might be a different drawing of $\mathbf{G}'$ with fewer crossings.

For each instance of this random selection we have the inequality (8.3), that is to say

$$\mathrm{cross}(\mathbf{G}') \geqslant \mathbf{E}' - 3\mathbf{V}'.$$

Certainly, then

$$\widetilde{\mathrm{cross}}(\mathbf{G}') \geqslant \mathbf{E}' - 3\mathbf{V}'.$$

We may take expectations of the three random variables appearing here and deduce, using linearity of expectation, that

$$\mathbb{E}\widetilde{\mathrm{cross}}(\mathbf{G}') \geqslant \mathbb{E}\mathbf{E}' - 3\mathbb{E}\mathbf{V}'. \tag{8.4}$$

However, it is easy to see that

$$\mathbb{E}\mathbf{V}' = pV \qquad \text{and} \qquad \mathbb{E}\mathbf{E}' = p^2 E$$

(since, for each edge in G, both endpoint vertices must be selected in order for it to be an edge in $\mathbf{G}'$), and

$$\mathbb{E}\widetilde{\mathrm{cross}}(\mathbf{G}') = p^4 \,\mathrm{cross}(G)$$

(since, for each crossing in G, all four endpoint vertices of the two edges involved must be selected in order for it to be a crossing in $\mathbf{G}'$).

Substituting into (8.4) gives

$$p^4 \,\mathrm{cross}(G) \geqslant p^2 E - 3pV.$$

We are free to choose any parameter $p \in [0,1]$ that we like. Choosing $p = 4V/E$ (noting that, by the hypothesis, $p \leqslant 1$) gives the desired bound after rearranging the terms. □

8.2. **The Szemerédi-Trotter theorem.** In this section we prove Theorem 8.1.

Proof of Theorem 8.1. First of all, note that Theorem 8.1 is trivial if $r \leqslant 7$ (say) since there are at most $\binom{m}{2}$ points lying on two or more lines.

Suppose henceforth that $r \geqslant 8$. Draw a graph G as follows. The vertices of G are the points P lying on at least r lines in L. Two vertices x, y are joined by an edge if and only if x, y are *consecutive* points *of* P on the same line in L. Denote by $n = |P|$ the number of vertices in this graph; this is the quantity we wish to bound.

Now observe that G comes with a natural drawing, that is to say the one induced by the lines in L; in this drawing, every edge is in fact represented by a straight line segment. Since two lines intersect in at most one point, the number of crossings in this drawing is at most $\binom{m}{2}$. Therefore

$$\mathrm{cross}(G) \leqslant \binom{m}{2} < m^2. \tag{8.5}$$

The number of edges E is at least $rn - m$. To see why, count the number of edges starting at v. Usually, v is adjacent to at least $2r$ other vertices. The exception is when v is one of the two endmost points (in either direction) on one of the lines in L, in which case we lose one adjacency. Summing over v gives at least $2rn - 2m$ pairs (v, w) with vw an edge, which of course double-counts the number of edges.

Now consider Proposition 8.3. Either we are in a position to apply this proposition, or we are not. If not, then $E < 4n$, so $rn - m < 4n$. Since $r \geqslant 8$, this implies that $rn/2 \leqslant m$, and so Theorem 8.1 holds in this case. Otherwise, $E \geqslant 4n$ and we may apply Proposition 8.3. This gives

$$m^2 \geqslant \frac{(rn-m)^3}{64n^2}. \tag{8.6}$$

If $n \leqslant 2m/r$ then again Theorem 8.1 holds. Otherwise, $rn - m \geqslant rn/2$ and so (8.6) becomes

$$m^2 \geqslant \frac{(rn/2)^3}{64n^2},$$

which immediately rearranges to $n \ll m^2/r^3$, and once again Theorem 8.1 holds. This concludes the proof. □

8.3. **Sum-product.** In this section we give Elekes's bound for the sum-product problem.

Theorem 8.4. *Suppose that $A \subset \mathbf{R}$ is a finite set of size n. Then $|A+A||A\cdot A| \gg n^{5/2}$. In particular, at least one of $A+A, A\cdot A$ has cardinality $\gg n^{5/4}$.*

Proof. It clearly suffices to handle the case $0 \notin A$ (otherwise remove 0 and apply the bound to the resulting set). Consider the set of points

$$P := \{(\frac{1}{a'}, -\frac{1}{aa'}) : a, a' \in A\},$$

and the set of lines

$$L := \{\{(x,y) \in \mathbf{R}^2 : ux + vy = 1\} : u \in A+A, v \in A \cdot A\}.$$

Observe that $|P| = n^2$, whilst the number $m = |L|$ of lines is $|A+A||A\cdot A|$.

The crucial observation is now that every point of P lies on at least n of these lines. Indeed, the point $(\frac{1}{a'}, -\frac{1}{aa'})$ lies on the line $ux + vy = 1$ when $u = a' + t$ and $v = at$, for every $t \in A$.

It follows from Szemerédi-Trotter that

$$n^2 \ll \frac{m}{n} + \frac{m^2}{n^3},$$

which implies that $m \gg n^{5/2}$. This is the desired result. □

9. Higher sum-product theorems

In this chapter, we will be considering higher-order sumsets and product sets of sets of integers. If $A \subset \mathbf{Z}$ is finite, and if $m \geqslant 1$ is an integer, we have already defined

$$mA := \{a_1 + \cdots + a_m : a_i \in A\}.$$

We now further define

$$A^{(m)} := \{a_1 \cdots a_m : a_i \in A\}.$$

Note that $2A = A + A$ and $A^{(2)} = A \cdot A$.

We showed in Theorem 8.4 that if $A \subset \mathbf{R}$ then either $2A$ or $A^{(2)}$ has size appreciably bigger than that of A, in fact size at least roughly $|A|^{5/4}$. In this section we will prove a more difficult result due to Bourgain and Chang, which asserts that if $A \subset \mathbf{N}$ then either mA or $A^{(m)}$ is *much* bigger than A, for large values of m. Here is the result we will prove.

Theorem 9.1. *Let $A \subset \mathbf{N}$. Then for any m either the m-fold sumset $|mA|$ or the m-fold product set $A^{(m)}$ has cardinality at least $|A|^{b(m)}$, where $b(m) \geqslant c \log m / \log\log m$.*

Note that for our proof it is important that A is a set of integers, though recently a corresponding result was shown for $A \subset \mathbf{R}$ by Mudgal [5] using the techniques of this course and additional deep inputs from diophantine analysis (the Subspace Theorem). Theorem 9.1 is due to Pálvölgyi and Zhelezov [10]; their proof is much easier than the original argument of Bourgain and Chang [2], and leads to a stronger bound. We will give a variant of the argument of [10] here using PFR, which was not available when the authors of [10] wrote their paper.

Remark. The original bound of Bourgain and Chang is on the order $b(m) \gg \log^{1/4} m$. The main point of these results is that $b(m) \to \infty$, which is a highly-nontrivial fact.

9.1. **Higher-order additive energies.** We begin by generalising the notion of additive energy, which we introduced in Section 3.

Definition 9.2. Let $k \geqslant 2$ be an integer. Given an additive set X, its additive $(2k)$-energy $E_{2k}(X)$ is the number of $(2k)$-tuples $(x_1, \dots, x_{2k}) \in X^{2k}$ such that $x_1 + \cdots + x_k = x_{k+1} + \cdots + x_{2k}$. More generally, if $X_1, \dots, X_{2k}$ are additive sets then we define $E(X_1, \dots, X_{2k})$ to be the number of solutions to $x_1 + \cdots + x_k = x_{k+1} + \cdots + x_{2k}$ with $x_i \in X_i$ for all i.

Thus $E_4(X)$ is the number of quadruples (x_1, x_2, x_3, x_4) such that $x_1 + x_2 = x_3 + x_4$, which is what we called simply the *additive energy* in Section 3, where we denoted it by $E(X)$.

We will need the following inequality.

Lemma 9.3. *Let $X_1, \dots, X_{2k} \subset \mathbf{Z}$ be finite sets. Then we have*

$$E(X_1, \dots, X_{2k}) \leqslant \prod_{i=1}^{2k} E_{2k}(X_i)^{1/2k}.$$

Proof. A quick proof of this may be given using the Fourier transform and Hölder's inequality. For this, observe that

$$E(X_1, \dots, X_{2k}) = \int_0^1 \widehat{1_{X_1}}(\theta) \cdots \widehat{1_{X_k}}(\theta) \overline{\widehat{1_{X_{k+1}}}(\theta)} \cdots \overline{\widehat{1_{X_{2k}}}(\theta)} d\theta,$$

where for any set

$$\widehat{1_X}(\theta) := \sum_n 1_X(n) e(-n\theta),$$

as in Section 1. The proof involves simply substituting the definition of the Fourier transform and using orthogonality, exactly as for the proof of (1.6).

Similarly (in fact, consequently) we have

$$E_{2k}(X_i) = \int_0^1 |\widehat{1_{X_i}}(\theta)|^{2k} d\theta.$$

The stated inequality is now a consequence of Hölder's inequality on the Fourier side, that is to say the inequality

$$\int_0^1 f_1 \cdots f_{2k} \leqslant \prod_{i=1}^{2k} (\int_0^1 |f_i|^{2k})^{1/2k}.$$

This concludes the proof. □

We will also need the following, which is essentially the higher-order version of Proposition 3.12, proved in the same way.

Lemma 9.4. *Let X be an additive set, and let $k \geqslant 2$ be an integer. Then*

$$|kX| \geqslant \frac{|X|^{2k}}{E_{2k}(X)}.$$

Proof. Write $r_k(n)$ for the number of tuples $(x_1, \dots, x_k) \in X^k$ with $x_1 + \cdots + x_k = n$. Then we have, by the Cauchy-Schwarz inequality,

$$|X|^{2k} = \big(\sum_{n \in kX} r_k(n) \big)^2 \leqslant |kX| \sum_n r_k(n)^2 = |kX| E_{2k}(X).$$

This concludes the proof. □

9.2. **A lemma of Chang.** If p is a prime and $m \in \mathbf{N}$, write $v_p(m)$ for the p-adic valuation of m, that is to say the exponent of the largest power of p dividing m. We have the following lemma of Mei-Chu Chang.

Lemma 9.5. *Let p be a prime, and suppose that $A \subset \mathbf{N}$ is a finite set. Let $A_i := \{n \in A : v_p(n) = i\}$. Then*

$$E_{2k}(A)^{1/k} \leqslant \binom{2k}{2} \sum_i E_{2k}(A_i)^{1/k}.$$

Proof. Since A is the disjoint union of the A_i, we have

$$E_{2k}(A) = \sum_{j_1,\dots,j_{2k}} E(A_{j_1}, \dots, A_{j_{2k}}).$$

However, not all of the terms here make any contribution. For a nonzero contribution we must have

$$p^{j_1}n_1 + \dots + p^{j_k}n_k = p^{j_{k+1}}n_{k+1} + \dots + p^{j_{2k}}n_{2k}$$

for some n_i coprime to p. Let $j = \min(j_1, \dots, j_{2k})$. Dividing through by p^j and considering congruences mod p, we see that there must be two i, i' with $j_i = j_{i'} = j$. Let us estimate the contribution in the case $\{i, i'\} = \{1, 2\}$; the other cases are essentially identical. This contribution is

$$\begin{aligned}\sum_{j,j_3,j_4,\dots,j_{2k}} E(A_j, A_j, A_{j_3}, \dots, A_{j_{2k}}) &= \sum_j E(A_j, A_j, A, \dots, A) \\ &\leqslant \sum_j E_{2k}(A_j)^{1/k} E_{2k}(A)^{(k-1)/k},\end{aligned}$$

where in the last step we used Lemma 9.3. Summing over the $\binom{2k}{2}$ choices of the pair $\{i, i'\}$ now gives

$$E_{2k}(A) \leqslant \binom{2k}{2} E_{2k}(A)^{(k-1)/k} \sum_j E_{2k}(A_j)^{1/k},$$

from which the lemma follows immediately. □

9.3. **The Bourgain-Chang theorem.** Suppose that $A \subset \mathbf{N}$ is a finite set. By the *multiplicative dimension* of A, we mean the dimension of the image of A under the map $v := (v_p)_{p\,\text{prime}} : \mathbf{N} \to \prod_p \mathbf{Z} \subset \prod_p \mathbf{Q}$. By *dimension*, we mean the dimension of the smallest affine subspace (translate of a vector subspace) of $\prod_p \mathbf{Q}$. Note that since A is a finite set, only finitely many primes are relevant here, so we can assume the image of v is finite-dimensional. We denote the multiplicative dimension by $\dim^{\times}(A)$.

Proposition 9.6. *Suppose that $A \subset \mathbf{N}$ is a set with multiplicative dimension at most D. Then*

$$E_{2k}(A)^{1/k} \leqslant \binom{2k}{2}^D |A|.$$

Proof. The result is almost immediate using Lemma 9.5 and induction on D, the result being trivial when $D = 0$ (in which case A is a singleton and $E_{2k}(A) = 1$). Otherwise, there is some prime p such that the image of the "coordinate map" $v_p : A \to \mathbf{Z} \subset \mathbf{Q}$ has size at least 2, and hence dimension 1; then the fibres this map all have dimension $D - 1$. These fibres, however, are precisely the A_i in the statement of Lemma 9.5. □

Now we turn to the main result, Theorem 9.1.

Proof of Theorem 9.1. At the expense of reducing c slightly, it suffices to handle the case when $m = 2^t$ is a sufficiently large power of two. Set $k := \lfloor \frac{t}{\log t} \rfloor$ and $b := \frac{\varepsilon t}{\log t}$ with $\varepsilon > 0$ some absolute constant to be specified later.

Suppose that

$$|A^{(2^t)}| \leqslant |A|^b. \tag{9.1}$$

Our aim is to show that

$$|2^t A| \geqslant |A|^b, \tag{9.2}$$

which will conclude the proof. The assumption (9.1) implies that

$$\prod_{i=0}^{t-1} \frac{|A^{(2^{i+1})}|}{|A^{(2^i)}|} \leqslant |A|^b,$$

so there is some $i \leqslant t-1$ such that

$$|A^{(2^{i+1})}| \leqslant K|A^{(2^i)}| \tag{9.3}$$

where $K = |A|^{b/t}$.

By the weak PFR over $\mathbf{Z}$, Theorem 7.1, there is a set $S \subset A^{(2^i)}$, $|S| \geqslant K^{-C}|A^{(2^i)}|$, with $\dim^{\times}(S) \leqslant C \log K$. Here, as before, $\dim^{\times}$ denotes the multiplicative dimension.

In the following argument, we will use the fact that if $X \subset \mathbf{N}$ has $|X \cdot X| \leqslant K|X|$ then $|X \cdot X^{-1}| \leqslant K^2|X|$, where $X^{-1} := \{x^{-1} : x \in X\}$. This follows from the Ruzsa triangle inequality Lemma 3.1 written multiplicatively, with $V = W = X$ and $U = X^{-1}$. (Note that $\mathbf{N}$ (with multiplication) is contained in the abelian group $\mathbf{Q}^{\times}$.)

Now we have $\sum_x |A \cap xS| = |A||S|$, and the sum is supported on $x \in AS^{-1}$. By the fact in the previous paragraph, the containment $S \subset A^{(2^i)}$ and (9.3), we have

$$|AS^{-1}| \leqslant |A^{(2^i)}(A^{(2^i)})^{-1}| \leqslant K^2|A^{(2^i)}|.$$

Therefore there is some x such that

$$|A \cap xS| \geqslant \frac{|A||S|}{|AS^{-1}|} \geqslant K^{-C-2}|A|.$$

Setting $A' := A \cap xS$, we therefore have

$$|A'| \geqslant K^{-C-2}|A| \tag{9.4}$$

and

$$\dim^{\times}(A') \leqslant \dim^{\times}(xS) = \dim^{\times}(S) \leqslant C \log K,$$

since multiplicative dimension is invariant under (multiplicative) translation.

By Proposition 9.6 and the dimension bound just proven,

$$E_{2k}(A') \leqslant \binom{2k}{2}^{Ck \log K} |A'|^k \leqslant K^{3Ck \log k}|A'|^k,$$

where in the second bound we used the crude bound $\log \binom{2k}{2} \leqslant \log(2k^2) < 3 \log k$ (since k is sufficiently large). By Lemma 9.4 and (9.4) (and, again, since k is sufficiently large) it follows that

$$|kA'| \geqslant \frac{|A'|^{2k}}{E_{2k}(A')} \geqslant K^{-3Ck \log k}|A'|^k \geqslant K^{-4Ck \log k}|A|^k.$$

Finally, we have

$$|2^t A| \geqslant |kA'| \geqslant K^{-4Ck \log k}|A|^k = |A|^{k - \frac{4Cbk \log k}{t}} \geqslant |A|^{k/2} > |A|^b$$

by the choice of b and k (if $\varepsilon < \min(\frac{1}{2}, \frac{1}{8C})$). This is (9.2), the bound we aimed to prove, so the proof is finished. □

Appendix A. Proof of Balog–Szemerédi–Gowers

A.1. **Paths of length 2.** The proof of the Balog-Szemerédi-Gowers theorem proceeds via the language of graph theory, establishing two lemmas of interest in their own right. The first, concerning paths of length 2, has the cleverer proof.

Lemma A.1. *Suppose that G is a bipartite graph on vertex set $V \cup W$, where $|V| = |W| = n$, and with αn^2 edges all of which join a vertex in V to one in W. Let $\eta > 0$ be a further parameter. Then there is a subset $V' \subseteq V$ with $|V'| \geqslant \alpha n/2$ such that between $(1-\eta)|V'|^2$ of the ordered pairs of points $(v_1, v_2) \in V' \times V'$ there are at least $\eta\alpha^2 n/2$ paths of length 2.*

Proof. If $x \in G$, write $N(x)$ for the neighbourhood of x in G, or in other words the set of vertices in G which are joined to x by an edge. Note that, since G is bipartite, $N(v) \subseteq W$ whenever $v \in V$ and $N(w) \subseteq V$ whenever $w \in W$.

Now by a double-counting argument, we have

$$\sum_{w\in W}\sum_{v\in V} 1_{vw\in E(G)} = \alpha n^2,$$

where $E(G)$ is of course the set of edges of G. Applying Cauchy-Schwarz to this gives

$$\sum_{w\in W}\sum_{v,v'\in V} 1_{vw\in E(G)}1_{v'w\in E(G)} \geqslant \alpha^2 n^3,$$

or in other words

$$\mathbb{E}_{v,v'\in V}|N(v)\cap N(v')| \geqslant \alpha^2 n. \tag{A.1}$$

This constitutes the rather basic observation that, on average, pairs (v, v') have many common neighbours. Now say that two vertices v and v' are *extremely unfriendly* if $|N(v) \cap N(v')| < \eta\alpha^2 n/2$, or in other words if there are fewer than $\eta\alpha^2 n/2$ paths of length two between v and v'. Write $S \subseteq V \times V$ for the set of extremely unfriendly pairs. Manifestly, from (A.1), we have

$$\mathbb{E}_{v,v'\in V}(\eta - 1_{(v,v')\in S})|N(v)\cap N(v')| \geqslant \eta\alpha^2 n/2.$$

This may be rewritten as

$$\mathbb{E}_{v,v'\in V}(\eta - 1_{(v,v')\in S})\sum_{w\in W} 1_{vw\in E(G)}1_{v'w\in E(G)} \geqslant \eta\alpha^2 n/2.$$

Turning the sum over W into an expectation (by dividing by $|W| = n$) and swapping the order of summation, this implies that

$$\mathbb{E}_{w\in W}\mathbb{E}_{v,v'\in V}(\eta - 1_{(v,v')\in S})1_{v,v'\in N(w)} \geqslant \eta\alpha^2/2.$$

In particular there is a choice of w such that

$$\mathbb{E}_{v,v'\in V}(\eta - 1_{(v,v')\in S})1_{v,v'\in N(w)} \geqslant \eta\alpha^2/2.$$

Simply the fact that this expectation is greater than zero tells us that at most a proportion η of the pairs $v, v' \in N(w)$ are extremely unfriendly. Furthermore (ignoring the term involving S completely) we have

$$\mathbb{E}_{v,v'\in V}1_{v,v'\in N(w)} \geqslant \alpha^2/2,$$

which implies that $|N(w)| \geqslant \alpha/\sqrt{2}$. Taking $V' := N(w)$, this proves the result. □

Remarks. This proof looks extremely slick at first sight. However when faced with the task of proving Lemma A.1 it is not hard to develop the feeling that one must somehow select a very "connected" subset of V. The way we have done this is essentially by picking a random vertex $w \in W$, and taking V' to be the neighbourhood $N(w)$ of w in V, though this was easier to manage by using expectations rather than starting with "pick $w \in W$ uniformly at random and consider

$N(w)$". This kind of technique seems to have been pioneered in this context by Gowers, and it is called "dependent random selection": one chooses something random (w in this case), then makes a deterministic choice based on it ($N(w)$).

A.2. **Paths of length 3.**

Lemma A.2. *Suppose that G is a bipartite graph on vertex set $V \cup W$, where $|V| = |W| = n$, and with αn^2 edges all of which join a vertex in V to one in W. Then there are subsets $V' \subseteq V$ and $W' \subseteq W$ with $|V'|, |W'| \geqslant c\alpha^C n$ such that between every pair $v' \in V'$ and $w' \in W'$ there are at least $c\alpha^C n^2$ paths of length 3 in G.*

Proof. Delete all edges emanating from vertices in V with degree less than $\alpha n/2$; this causes the deletion of at most $\alpha n^2/2$ edges in total, so at least $\alpha n^2/2$ remain. From now on if we speak of an *edge* we mean one of these edges. Let $\eta > 0$ be a parameter to be chosen later. Using the preceding lemma, we may select a set $V' \subseteq V$ with $|V'| \geqslant \alpha n/4$ such that a proportion $1 - \eta$ of the pairs of vertices in V' have at least $\eta\alpha^2 n/8$ common neighbours in W.

All vertices in V' have degree 0 or else degree at least $\alpha n/2$, but it is conceivably the case that some do have degree 0. However if $\eta < 1/4$ then clearly no more than half of them do. Thus we may pass to a set $V'' \subseteq V'$, $|V''| \geqslant \alpha n/8$, such that every vertex in V'' has degree at least $\alpha n/2$ and still such that a proportion $1 - \eta$ of the pairs of vertices in V'' have at least $\eta\alpha^2 n/8$ common neighbours in W.

Now let us focus on W. Look at all the edges from V'' into W: since each vertex in V'' has degree at least $\alpha n/2$, and $|V''| \geqslant \alpha n/8$, there are at least $\alpha^2 n^2/16$ of these. It follows that there is some set $W' \subseteq W$, $|W'| \geqslant \alpha^2 n/32$, such that each $w \in W'$ has at least $\alpha^2 n/32$ neighbours in V''.

Before concluding, let us jump back over to the other side and effect one final refinement of V''. Say that a vertex $v \in V''$ is *sociable* if there is a proportion at least $1 - 2\eta$ of the other vertices $v' \in V''$ are such that v and v' have at least $\eta\alpha^2 n/8$ common neighbours. Then at least half the vertices of V'' are sociable: call this set V''', so that $|V'''| \geqslant \alpha n/16$.

We now claim that for any $x \in V'''$ and $y \in W'$ there are many paths of length three between x and y (in the original graph G). Indeed by the choice of W' there must be at least $\alpha^2 n/32$ elements of V'' adjacent to y. There must also be at least $(1 - 2\eta)|V''|$ vertices of V'' which have at least $\eta\alpha^2 n/8$ common neighbours with x. Provided that $\alpha^2 n/32 \geqslant 3\eta|V''|$, which will be the case if $\eta \leqslant \alpha^2/96$, these two sets intersect in a set $\tilde{V} \subseteq V''$ of size at least $\eta|V''|$. Thus each element z of $\tilde{V}$ is adjacent to y, and has $\eta\alpha^2 n/8$ common neighbours with x. This clearly leads to at least $\eta^2\alpha^2|V''|n/8$ paths of length three between x and y.

The only constraints on η were that $\eta \leqslant 1/4$ and that $\eta \leqslant \alpha^2/96$. The latter is clearly the more severe constraint, so set $\eta := \alpha^2/96$. The lemma is proven. □

A.3. **Proof of Balog-Szemerédi-Gowers.** In this section we deduce Theorem 3.13 from the paths of length 3 lemma, Lemma A.2. It is particularly important to remember during this proof that the constant C may change from line to line.

Proof of Theorem 3.13. For the majority of the proof we handle the two-sets case (i) and the one-set case (ii) at the same time, taking $A = B$ in the latter case.

Suppose then that A, B are two sets in some abelian group G and that $\omega[A, B] \geqslant 1/K$. This means that there are at least $|A|^{3/2}|B|^{3/2}/K$ solutions to $a_1 - b_1 = a_2 - b_2$. Note that the number of solutions to this equation is at most $|A|^2|B|$, since once a_1, b_1 and a_2 are specified b_2 is uniquely determined. Therefore $|B| \leqslant K^2|A|$, and similarly $|A| \leqslant K^2|B|$.

Write $s(x)$ for the number of pairs $(a, b) \in A \times B$ with $a - b = x$. Thus we have

$$\sum_x s(x)^2 \geqslant |A|^{3/2}|B|^{3/2}/K,$$

whilst by double-counting pairs $(a, b) \in A \times B$ we have

$$\sum_x s(x) = |A||B|.$$

We claim there are at least $|A|^{1/2}|B|^{1/2}/2K$ "popular" values of x for which $s(x) \geqslant |A|^{1/2}|B|^{1/2}/2K$. To see this, let Δ denote the set of these popular x. Then

$$\sum_{x \notin \Delta} s(x)^2 \leqslant \frac{1}{2K}|A|^{1/2}|B|^{1/2} \sum_x s(x) = |A|^{3/2}|B|^{3/2}/2K,$$

so

$$\sum_{x \in \Delta} s(x)^2 \geqslant |A|^{3/2}|B|^{3/2}/2K.$$

However, since $s(x) \leqslant \min(|A|, |B|) \leqslant |A|^{1/2}|B|^{1/2}$ for every x,

$$\sum_{x \in \Delta} s(x)^2 \leqslant |\Delta||A||B|.$$

The claim follows.

Note also, for use below, that

$$|\Delta| \leqslant 2K|A|^{1/2}|B|^{1/2}, \tag{A.2}$$

a bound which follows straightforwardly by double-counting pairs $(a, b) \in A \times B$.

Define a bipartite graph G on vertex set $A \cup B$ by joining $a \in A$ to $b \in B$ by an edge if $a - b$ is a popular difference in the above sense, that is to say if and only if $a - b \in \Delta$. Then G has at least $|A||B|/4K^2$ edges. Let $n = \max(|A|, |B|)$, and "pad out" the smaller vertex class of G to obtain a new graph having n vertices in each class. Recalling that $K^{-2} \leqslant |A|/|B| \leqslant K^2$, this graph has at least $n^2/4K^4$ edges.

Applying Lemma A.2, we may locate sets $A' \subseteq A$ and $B' \subseteq B$ with $|A'| \gg K^{-C}|A|$, $|B'| \gg K^{-C}|B|$ and such that for every $a' \in A'$ and $b' \in B'$ there are $\gg K^{-C}n^2$ paths of length 3 in G between a' and b'. This, of course, means that there $\gg K^{-C}n^2$ choices of $a'' \in A$ and $b'' \in B$ such that all three of $a' - b''$, $a'' - b''$ and $a'' - b'$ lie in Δ.

Noting that $a' - b' = (a' - b'') - (a'' - b'') + (a'' - b')$, it follows that for all $a' \in A'$ and $b' \in B'$ the difference $a' - b'$ can be written in $\gg K^{-C}n^2$ ways as $x - y + z$, where $x, y, z \in \Delta$. These are genuinely distinct representations, since it is easy to recover a'' and b'' from knowledge of a', b', x, y and z. However, by (A.2), the number of popular differences is bounded *above* by $2K|A|^{1/2}|B|^{1/2} \ll Kn$. It follows that

$$|A' - B'| \cdot K^{-C}n^2 \ll (Kn)^3,$$

which of course implies that

$$|A' - B'| \ll K^C n. \tag{A.3}$$

To finish the argument, we consider parts (i) and (ii) of Theorem 3.13 separately.

In case (i), applying (A.3) and Corollary 3.9 together with the lower bounds $|A'|, |B'| \geqslant K^{-2}n$ gives the desired upper bound $|A' + B'| \ll K^C n \ll K^C|A'|^{1/2}|B'|^{1/2}$.

In case (ii), we first apply the Ruzsa triangle inequality with $U = B'$, $V = W = A'$ to conclude from (A.3) that $|A' - A'| \ll K^C n$. From this, it follows using Corollary 3.9 that $|A' + A'| \ll K^C n$. □

Appendix B. Entropic Balog-Szemerédi-Gowers

The material in this appendix is not examinable.

Lemma B.1 (Entropic BSG)**.** *Let (A, B) be a G^2-valued random variable, and set $Z \coloneqq A + B$. Then*

$$\sum_z p_Z(z)\mathrm{d}[(A|Z = z); (B|Z = z)] \leqslant 3\mathbf{I}[A : B] + 2\mathbf{H}[Z] - \mathbf{H}[A] - \mathbf{H}[B]. \tag{B.1}$$

We stress that the quantity $2\mathbf{H}[Z] - \mathbf{H}[A] - \mathbf{H}[B]$ is *not* the same as $2\mathrm{d}[A; B]$, because (A, B) are given a joint distribution which may not be independent. In particular, $\mathbf{H}[Z] = \mathbf{H}[A + B]$ may not match the entropy of a sum of independent copies of A and B.

Proof. In the proof we will need the notion of *conditionally independent trials* of a pair of random variables (X, Y) (not necessarily independent). We say that X_1, X_2 are conditionally independent trials of X relative to Y by declaring $(X_1|Y = y)$ and $(X_2|Y = y)$ to be independent copies of $(X|Y = y)$ for all y in the range of Y. We then have

$$\mathbf{H}[(X_1|Y = y), (X_2|Y = y)] = 2\mathbf{H}[X|Y = y]$$

for all y, which upon summing over y (weighted by $p_Y(y)$) gives

$$\mathbf{H}[X_1, X_2|Y] = 2\mathbf{H}[X|Y]$$

and hence

$$\begin{aligned}\mathbf{H}[X_1, X_2, Y] = \mathbf{H}[X_1, X_2|Y] + \mathbf{H}[Y] &= 2\mathbf{H}[X|Y] + \mathbf{H}[Y] \\ &= 2\mathbf{H}[X, Y] - \mathbf{H}[Y].\end{aligned} \tag{B.2}$$

Note also that the marginal distributions of (X_1, Y) and (X_2, Y) each match the original distribution (X, Y).

Turning to the proof of Lemma B.1 itself, let (A_1, B_1) and (A_2, B_2) be conditionally independent trials of (A, B) relative to Z, thus (A_1, B_1) and (A_2, B_2) are coupled through the random variable $A_1 + B_1 = A_2 + B_2$, which by abuse of notation we shall also call Z.

Observe that the left-hand side of (B.1) is

$$\mathbf{H}[A_1 - B_2|Z] - \tfrac{1}{2}\mathbf{H}[A_1|Z] - \tfrac{1}{2}\mathbf{H}[B_2|Z]. \tag{B.3}$$

since, crucially, $(A_1|Z = z)$ and $(B_2|Z = z)$ are independent for all z.

Applying submodularity (5.8) gives

$$\begin{aligned}&\mathbf{H}[A_1 - B_2] + \mathbf{H}[A_1 - B_2, A_1, B_1] \\ &\qquad \leqslant \mathbf{H}[A_1 - B_2, A_1] + \mathbf{H}[A_1 - B_2, B_1].\end{aligned} \tag{B.4}$$

We estimate the second, third and fourth terms appearing here. First note that, by (B.2) (noting that the tuple $(A_1 - B_2, A_1, B_1)$ determines the tuple (A_1, A_2, B_1, B_2) since $A_1 + B_1 = A_2 + B_2$)

$$\mathbf{H}[A_1 - B_2, A_1, B_1] = \mathbf{H}[A_1, B_1, A_2, B_2] = 2\mathbf{H}[A, B] - \mathbf{H}[Z]. \tag{B.5}$$

Next observe that

$$\mathbf{H}[A_1 - B_2, A_1] = \mathbf{H}[A_1, B_2] \leqslant \mathbf{H}[A] + \mathbf{H}[B]. \tag{B.6}$$

Finally, we have

$$\mathbf{H}[A_1 - B_2, B_1] = \mathbf{H}[A_2 - B_1, B_1] = \mathbf{H}[A_2, B_1] \leqslant \mathbf{H}[A] - \mathbf{H}[B]. \tag{B.7}$$

Substituting (B.5), (B.6) and (B.7) into (B.4) yields

$$\mathbf{H}[A_1 - B_2] \leqslant 2\mathbf{I}[A : B] + \mathbf{H}[Z]$$

and so by (5.6)

$$\mathbf{H}[A_1 - B_2|Z] \leqslant 2\mathbf{I}[A : B] + \mathbf{H}[Z].$$

Since

$$\begin{aligned}\mathbf{H}[A_1|Z] &= \mathbf{H}[A_1, A_1 + B_1] - \mathbf{H}[Z] \\ &= \mathbf{H}[A, B] - \mathbf{H}[Z] \\ &= \mathbf{H}[A] + \mathbf{H}[B] - \mathbf{I}[A : B] - \mathbf{H}[Z]\end{aligned}$$

and similarly for $\mathbf{H}[B_2|Z]$, we see that (B.3) is bounded by $3\mathbf{I}[A : B] + 2\mathbf{H}[Z] - \mathbf{H}[A] - \mathbf{H}[B]$ as claimed. □

Appendix C. Geometry of numbers

The material in this appendix is not examinable.

The main goal of this section is to prove Minkowski's second theorem. First we briefly go over some standard properties of the determinant of a lattice.

Lemma C.1. *If $q \in \mathbf{N}$ then $\det(q\mathbf{Z}^d) = q^d$. If Λ, Λ' are two lattices with $\Lambda' \subset \Lambda$, then $\det(\Lambda')/\det(\Lambda) = [\Lambda : \Lambda']$, where the latter quantity is the index of Λ' as a subgroup of Λ, that is to say the number of cosets of Λ' needed to cover Λ.*

Now let us recall the statement of Minkowski's Second theorem, and let us also state Minkowski's *first* theorem. In both of these results, $K \subset \mathbf{R}^d$ is a centrally symmetric convex body, and $\Lambda \subset \mathbf{R}^d$ a lattice. The successive minima of K with respect to Λ are $\lambda_1, \dots, \lambda_d$.

Theorem C.2 (Minkowski I)**.** *Suppose that $\mathrm{vol}(K) > 2^d \det(\Lambda)$. Then K contains a nonzero point of Λ.*

Theorem C.3 (Minkowski II)**.** *We have $\lambda_1 \cdots \lambda_d \,\mathrm{vol}(K) \leqslant 2^d \det(\Lambda)$.*

Let us remark that Minkwoski I is a consequence of Minkowski II. To see this, note that if $\mathrm{vol}(K) > 2^d \det(\Lambda)$ then Minkowski II implies that $\lambda_1 \cdots \lambda_d < 1$. Since $\lambda_1 \leqslant \cdots \lambda_d$, this implies that $\lambda_1 < 1$. By the definition of λ_1, it follows that K contains at least one nonzero point of Λ.

Minkowski I is a very straightforward consequence of the following result, *Blichfeldt's lemma*, which is also an ingredient in the proof of Minkowski II.

Lemma C.4 (Blichfeldt's lemma)**.** *Suppose that $K \subset \mathbf{R}^d$, and suppose that $\mathrm{vol}(K) > \det(\Lambda)$. Then there are two distinct points $\mathbf{x}, \mathbf{y} \in K$ with $\mathbf{x} - \mathbf{y} \in \Lambda$.*

Remark. Note that here K is not required to be either centrally symmetric or convex.

Proof. By considering the sets $K \cap B(0, R)$, as $R \to \infty$, whose volumes tend to that of K, we may assume that K lies inside some ball $B(0, R)$. Now let us suppose that the conclusion is false: then no translate of K contains two points of Λ, or in other words

$$\sum_{\mathbf{x}} 1_K(\mathbf{x} - \mathbf{t}) 1_\Lambda(\mathbf{x}) \leqslant 1$$

for all $\mathbf{t} \in \mathbf{R}^d$. Let R' be much bigger than R, and average this last inequality over $\mathbf{t}$ lying in the ball $B(0, R')$ to obtain

$$\sum_x 1_\Lambda(\mathbf{x}) \big(\frac{1}{\mathrm{vol}(B(0, R'))} \int_{B(0,R')} 1_K(\mathbf{x} - \mathbf{t}) d\mathbf{t} \big) \leqslant 1.$$

Since $K \subset B(0, R)$, the inner integral equals $\mathrm{vol}(K)$ if $\|x\| \leqslant R' - R$, and therefore

$$\sum_{\mathbf{x}} 1_\Lambda(\mathbf{x}) 1_{B(0,R'-R)}(\mathbf{x}) d\mathbf{x} \leqslant \frac{\mathrm{vol}(B(0, R')}{\mathrm{vol}(K)},$$

and hence

$$\frac{1}{\mathrm{vol}(B(0,R'-R))}\sum_{\mathbf{x}} 1_{\Lambda}(\mathbf{x})1_{B(0,R'-R)}(\mathbf{x})d\mathbf{x} \leqslant \frac{\mathrm{vol}(B(0,R')}{\mathrm{vol}(B(0,R'-R))}\cdot\frac{1}{\mathrm{vol}(K)}. \tag{C.1}$$

However it is "clear" by tiling with fundamental parallelepipeds that

$$\lim_{r\to\infty}\frac{1}{\mathrm{vol}(B(0,r))}\sum_{\mathbf{x}} 1_{\Lambda}(\mathbf{x})1_{B(0,r)}(\mathbf{x}) = \frac{1}{\det(\Lambda)},$$

and moreover

$$\lim_{R'\to\infty}\frac{\mathrm{vol}(B(0,R')}{\mathrm{vol}(B(0,R'-R))} = 1.$$

Comparing with (C.1) immediately leads to

$$\frac{1}{\det(\Lambda)} \leqslant \frac{1}{\mathrm{vol}(K)},$$

contrary to assumption. □

Although we will not formally need it in what follows, let us pause to give the simple deduction of Minkowski I.

Proof of Minkowski I. By Blichfeldt's lemma, the set $\frac{1}{2}K = \{\frac{1}{2}\mathbf{x} : \mathbf{x}\in\mathbf{R}^d\}$ contains two distinct points of Λ; thus there are $\mathbf{x},\mathbf{y}\in K$ with $\frac{1}{2}(\mathbf{x}-\mathbf{y})\in\Lambda$. However, since K is convex and centrally symmetric we have $\frac{1}{2}(\mathbf{x}-\mathbf{y})\in K$. □

Now we turn to the proof of Minkowski II.

Proof of Minkowski II. It is technically convenient to assume that K is *open*; this we may do by passing from K to the interior K°. Take a directional basis $\mathbf{b}_1,\dots,\mathbf{b}_d$ for Λ with respect to K. Since K is open, $\lambda_k K\cap\Lambda$ is spanned (over $\mathbf{R}$) by the vectors $\mathbf{b}_1,\dots,\mathbf{b}_{k-1}$. Indeed if it were not then we could choose some further linearly independent vector $\mathbf{b}\in\lambda_k K\cap\Lambda$, and by the openness of K this would in fact lie in $(\lambda_k-\varepsilon)K\cap\Lambda$ for some $\varepsilon>0$, contrary to the definition of λ_k.

Write each given $\mathbf{x}$ in coordinates relative to the basis vectors $\mathbf{b}_i$ as $x_1\mathbf{b}_1+\cdots+x_d\mathbf{b}_d$. We now define some rather unusual maps $\phi_j : K\to K$, by mapping $\mathbf{x}\in K$ to the centre of gravity of the slice of K which contains $\mathbf{x}$ and is parallel to the subspace spanned by $\mathbf{b}_1,\cdots,\mathbf{b}_{j-1}$ (for $j=1$, $\phi_1(\mathbf{x})=\mathbf{x}$). Next, we define a map $\phi : K\to\mathbf{R}^d$ by

$$\phi(\mathbf{x}) := \sum_{j=1}^{d}(\lambda_j-\lambda_{j-1})\phi_j(\mathbf{x}),$$

where we are operating with the convention that $\lambda_0=0$. Let us make a few further observations concerning the ϕ_j and ϕ. In coordinates we have $\phi_j(\mathbf{x})=\sum_i c_{ij}(\mathbf{x})\mathbf{b}_i$, where $c_{ij}(\mathbf{x})=x_i$ for $i\geqslant j$, and $c_{ij}(\mathbf{x})$ depends only on $x_j,\cdots,x_d$ for $i<j$. It follows that

$$\phi(\mathbf{x}) = \sum_{i=1}^{d}\mathbf{b}_i(\lambda_i x_i+\psi_j(x_{i+1},\cdots,x_d))$$

for certain continuous functions ψ_j. It follows easily that

$$\mathrm{vol}(\phi(K)) = \lambda_1\cdots\lambda_d\,\mathrm{vol}(K), \tag{C.2}$$

the Jacobian of the transformation $x_i'=\lambda_i x_i+\psi_i(x_{i+1},\dots,x_d)$ being $\lambda_1\cdots\lambda_d$.

Suppose, as a hypothesis for contradiction, that $\lambda_1\cdots\lambda_d\,\mathrm{vol}(K) > 2^d\det(\Lambda)$. By Blichfeldt's lemma and (C.2), this means that $\phi(K)$ contains two elements $\phi(\mathbf{x})$ and $\phi(\mathbf{y})$ which differ by an element of $2\cdot\Lambda=\{2\lambda : \lambda\in\Lambda\}$, and this means that $\frac{1}{2}(\phi(\mathbf{x})-\phi(\mathbf{y}))\in\Lambda$. Write $\mathbf{x}=\sum_i x_i\mathbf{b}_i$ and

$\mathbf{y} = \sum_i y_i \mathbf{b}_i$, and suppose that k is the largest index such that $x_k \neq y_k$. Then we have $\phi_i(\mathbf{x}) = \phi_i(\mathbf{y})$ for $i > k$, so that

$$\begin{aligned}\frac{\phi(\mathbf{x}) - \phi(\mathbf{y})}{2} &= \sum_{j=1}^{d} (\lambda_j - \lambda_{j-1}) \Big(\frac{\phi_j(\mathbf{x}) - \phi_j(\mathbf{y})}{2} \Big) \\ &= \sum_{j=1}^{k} (\lambda_j - \lambda_{j-1}) \Big(\frac{\phi_j(\mathbf{x}) - \phi_j(\mathbf{y})}{2} \Big).\end{aligned}$$

This has two consequences. First of all the convexity of K implies that $\frac{1}{2}(\phi_j(\mathbf{x}) - \phi_j(\mathbf{y})) \in K$ for all j, and hence (again by convexity) $\frac{1}{2}(\phi(\mathbf{x}) - \phi(\mathbf{y})) \in \lambda_k K$. Secondly we may easily evaluate the coefficient of $\mathbf{b}_k$ when $\frac{1}{2}(\phi(\mathbf{x}) - \phi(\mathbf{y}))$ is written in terms of our directional basis: it is exactly $\lambda_k(x_k - y_k)/2$. In particular this is nonzero, which means that $\frac{1}{2}(\phi(\mathbf{x}) - \phi(\mathbf{y}))$ lies in Λ and $\lambda_k K$, but not in the span of $\mathbf{b}_1, \cdots, \mathbf{b}_{k-1}$. This is contrary to the observation made at the start of the proof. □

References

[1] Noga Alon and Joel H. Spencer, *The probabilistic method*, fourth ed., Wiley Series in Discrete Mathematics and Optimization, John Wiley & Sons, Inc., Hoboken, NJ, 2016.

[2] Jean Bourgain and Mei-Chu Chang, *On the size of k-fold sum and product sets of integers*, J. Amer. Math. Soc. **17** (2004), 473–497.

[3] W. T. Gowers, Ben Green, Freddie Manners, and Terence Tao, *On a conjecture of Marton*, arXiv:2311.05762.

[4] Ben Green, Freddie Manners, and Terence Tao, *Sumsets and entropy revisited*, Random Structures Algorithms **66** (2025), Paper No. e21252, 33.

[5] Akshat Mudgal, *An Elekes-Rónyai theorem for sets with few products*, arXiv:2308.04191.

[6] Melvyn B. Nathanson, *Additive number theory*, Graduate Texts in Mathematics, vol. 164, Springer-Verlag, New York, 1996, The classical bases.

[7] Melvyn B. Nathanson, *Additive number theory*, Graduate Texts in Mathematics, vol. 165, Springer-Verlag, New York, 1996, Inverse problems and the geometry of sumsets.

[8] Terence Tao and Van H. Vu, *Additive combinatorics*, paperback ed., Cambridge Studies in Advanced Mathematics, vol. 105, Cambridge University Press, Cambridge, 2010.

[9] Yufei Zhao, *Graph theory and additive combinatorics—exploring structure and randomness*, Cambridge University Press, Cambridge, 2023.

[10] Dmitrii Zhelezov and Dömötör Pálvölgyi, *Query complexity and the polynomial Freiman-Ruzsa conjecture*, Adv. Math. **392** (2021), Paper No. 108043, 18.