Structure Theory of Set Addition

Notes by B. J. Green[1]

ICMS Instructional Conference in Combinatorial Aspects of Mathematical Analysis,
Edinburgh March 25 – April 5 2002.

1 Lecture 1: Plünnecke's Inequalities

1.1 Introduction

The object of these notes is to explain a recent proof by Ruzsa of a famous result of Freiman, some significant modifications of Ruzsa's proof due to Chang, and all the background material necessary to understand these arguments.

Freiman's theorem concerns the structure of sets with small sumset. Let A be a subset of an abelian group G, and define the sumset $A + A$ to be the set of all pairwise sums $a + a'$, where a, a' are (not necessarily distinct) elements of A. If $|A| = n$ then $|A + A| \geq n$, and equality can occur (for example if A is a subgroup of G). In the other direction we have $|A + A| \leq n(n + 1)/2$, and equality can occur here too, for example when $G = \mathbb{Z}$ and $A = \{1, 3, 3^2, \dots, 3^{n-1}\}$. It is easy to construct similar examples of sets with large sumset, but rather harder to find examples with $A + A$ small. Let us think more carefully about this problem in the special case $G = \mathbb{Z}$.

Proposition 1 *Let $A \subseteq \mathbb{Z}$ have size n. Then $|A + A| \geq 2n - 1$, with equality if and only if A is an arithmetic progression of length n.*

Proof. Write $A = \{a_1, \dots, a_n\}$ where $a_1 < a_2 < \cdots < a_n$. Then we have

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

which amounts to an exhibition of $2n - 1$ distinct elements of $A + A$. There are other ways of exhibiting $2n - 1$ distinct elements of $A + A$; for any $1 \leq i \leq n$ we have

$$a_1 + a_1 < \cdots < a_1 + a_i < \cdots < a_i + a_i < \cdots < a_i + a_n < \cdots < a_n + a_n.$$

If $|A + A| = 2n - 1$, however, these two orderings must coincide exactly for any i. Thus in particular we have $a_2 + a_i = a_1 + a_{i+1}$, and it is easy to see that this forces A to be an arithmetic progression. □

[1] Trinity College, Cambridge CB2 1TQ. I would be most grateful to receive comments and corrections, which may be emailed to me at bjg23@hermes.cam.ac.uk.

It is easy to exhibit rather less structured sets for which $|A+A| \leq 4n$, say; simply take an arithmetic progression of length $2n$ and select any n elements from it. Such sets are still covered rather economically by an arithmetic progression. There are, however, examples of sets with small sumset which are not of this form. Let $x_0, x_1, \dots, x_d \in \mathbb{Z}$ and let $m_1, \dots, m_d$ be positive integers. The set

$$P = \left\{ x_0 + \sum_{j=1}^{d} \lambda_j x_j \;\middle|\; 0 \leq \lambda_j \leq m_j - 1 \right\}$$

is said to be a d-dimensional progression. We say that P is *proper* if $|P| = m_1 m_2 \dots m_d$, that is to say if all of the sums comprising P are distinct. If P is proper then it is easy to confirm that $|P+P| \leq 2^d |P|$.

Freiman's beautiful and deep theorem states that these are essentially the only examples of subsets of $\mathbb{Z}$ with small sumset. The qualitative form of his result is the following.

Theorem 2 (Freiman,[4]) *Let $A \subseteq \mathbb{Z}$ have cardinality n, and suppose that $|A+A| \leq C|A|$. Then A is contained in a proper d-dimensional progression P of size at most Kn, where d and K depend only on C.*

1.2 Plünnecke's Inequalities

If k and l are positive integers then we may generalise the concept of a sumset as follows. Let A be a subset of an abelian group G, and write $kA - lA$ for the set consisting of all elements of G of the form $a_1 + \dots + a_k - a_1' - \dots - a_l'$ (we will also use such notations as $kA + lB$, whose meanings should be obvious).

Theorem 3 (Plünnecke – Ruzsa) *Suppose that $|A+A| \leq C|A|$. Then $|kA - lA| \leq C^{k+l}|A|$ for any k, l.*

Plünnecke proved some results which are at least in a similar spirit to this in several papers and in the monograph [12]. As well as being in German, these papers suffer from a surfeit of notation and I was unable to make much headway with them. Thus the worker in this area should be grateful to Ruzsa [15] for rediscovering and simplifying Plünnecke's work, so that we can give the polished treatment that follows.

We will deduce Plünnecke's inequalities from a rather different looking result. Before stating this we need to make some definitions. A *Plünnecke Graph* of level h is a directed graph$G = (V(G), E(G))$ on some vertex set $V_0 \cup V_1 \cup \dots \cup V_h$ satisfying the following properties:

(i) All edges in $E(G)$ are of the form (v, v') where $v \in V_i$ and $v' \in V_{i+1}$ for some $0 \leq i \leq h-1$;
(ii) (Forward splitting of paths) Let $0 \leq i \leq h-2$ and suppose that $u \in V_i$, $v \in V_{i+1}$ and $w_1, \dots, w_k \in V_{i+2}$ are such that (u, v) and all of the (v, w_j) are edges of G. Then there are

distinct $v_1, \dots, v_k \in V_{i+1}$ such that all of the (u, v_j) and the (v_j, w_j) are edges of G.
(iii) (Backward splitting of paths) Let $0 \leq i \leq h - 2$ and suppose that $u_1, \dots, u_k \in V_i$, $v \in V_{i+1}$ and $w \in V_{i+2}$ are such that (v, w) and all of the (u_j, v) are edges of G. Then there are distinct $v_1, \dots, v_k \in V_{i+1}$ such that all of the (u_j, v_j) and the (v_j, w) are edges of G.

It goes without saying that the reader is advised to draw a picture representing properties (ii) and (iii), whereupon their meaning will be clarified enourmously. Now if $X \subseteq V_0$ we write $\mathrm{im}_i(X)$ for the set of all vertices in V_i which can be reached by a path starting from some $x \in X$. The ith *magnification ratio* of G, $D_i(G)$, is defined by

$$D_i(G) \;=\; \inf_{X \subseteq V_0, X \neq \emptyset} \frac{|\mathrm{im}_i(X)|}{|X|}.$$

Proposition 4 (Plünnecke) *Let G be a Plünnecke graph of level $h \geq 2$. Then we have the inequalities*

$$D_1 \;\geq\; D_2^{1/2} \;\geq\; D_3^{1/3} \;\geq\; D_h^{1/h}.$$

The deduction of Theorem 3 from Proposition 4 is a relatively simple matter which, furthermore, furnishes us with an example of a Plünnecke graph which it may be useful to have in mind. The key step is the following.

Proposition 5 *Let A, B be subsets of an abelian group with $|A + hB| \leq C|A|$. Then, for any $h' \geq h$, there is a set $A' \subseteq A$ with $|A' + h'B| \leq C^{h'/h}|A'|$.*

Proof. Define a directed graph as follows. Set $V_i = A + iB$, and join $v \in V_i$ to $v' \in V_{i+1}$ precisely if $v' - v \in B$. It is very easy to check that the graph so defined is Plünnecke; we denote it by $\mathrm{Plün}(A, B)$. The hth magnification ratio, D_h, is at most C because

$$\inf_{Z \subseteq A} \frac{\mathrm{im}_h(Z)}{|Z|} \;\leq\; \frac{\mathrm{im}_h(A)}{|A|} \;\leq\; \frac{|A + hB|}{|A|} \;\leq\; C.$$

It follows from Proposition 4 that $D_{h'} \leq C^{h'/h}$ for any $h' \geq h$, which is equivalent to the statement of the proposition. □

It follows immediately, taking $h = 1$, that if $|A + A| \leq C|A|$ then $|kA| \leq C^k|A|$ for any $k \geq 2$. To deduce the full strength of Theorem 3 we need a further small lemma.

Lemma 6 *Let U, V, W be subsets of an abelian group. Then we have*

$$|U||V - W| \;\leq\; |U + V||U + W|.$$

Proof. For any $d \in V - W$ fix $v(d) \in V$, $w(d) \in W$ with $v(d) - w(d) = d$. We define a map from $U \times (V - W)$ to $(U + V) \times (U + W)$ by sending (u, d) to $(u + v(d), u + w(d))$. It is easy to check that this is injective. □

We may now complete the proof of Theorem 3. Suppose that $|A+A| \leq C|A|$, and suppose without loss of generality that $l \geq k$. We may apply Proposition 5 twice to get sets $A'' \subseteq A' \subseteq A$ satisfying

$$|A' + kA| \;\leq\; C^k|A'|$$

and

$$|A'' + lA| \;\leq\; C^l|A''|.$$

Lemma 6 then gives

$$\begin{aligned} |A''||kA - lA| &\leq |A'' + kA||A'' + lA| \\ &\leq |A' + kA||A'' + lA| \\ &\leq C^{k+l}|A'||A''| \\ &\leq C^{k+l}|A||A''|. \end{aligned}$$

Cancelling the common factor of $|A''|$ completes the deduction of Theorem 3 from Proposition 4.

We turn now to the proof of Proposition 4. This proposition certainly implies that if $D_h \geq 1$ then $D_i \geq 1$ for all $i \leq h$. It turns out that, with rather more effort, one can reverse this implication. Therefore we begin by proving

Proposition 7 *Let G be a Plünnecke graph of level h, and suppose that $D_h \geq 1$. Then $D_i \geq 1$ for all i.*

Proof. Let $V(G) = V_0 \cup \cdots \cup V_h$. We will show that if $D_h \geq 1$ then there are $|V_0|$ vertex-disjoint paths from V_0 to V_h, from which it will be immediate that $D_i \geq 1$ for all i. We will achieve this by invoking Menger's Theorem from graph theory (see, for example, [2]). This theorem tells us that the maximum number of vertex-disjoint paths is exactly equal to the size of the largest *separating set* in G, that is to say the smallest number of vertices we can choose so that every path contains at least one of them.

Let $S \subseteq V(G)$ be a separating set of minimal size m, and suppose in addition that S is as "bottom heavy" as possible, by which we mean that the sum

$$\sum_{i=0}^{h} i|S \cap V_i|$$

is minimal. Suppose, for a contradiction, that $m < |V_0|$. Our first claim is that S meets $V_1 \cup \cdots \cup V_{h-1}$. Indeed, suppose that $S = X \cup Y$, where $X \subseteq V_0$ and $Y \subseteq V_h$. S meets every path from V_0 to V_h and so Y must meet every path starting in $V_0 \backslash X$, that is $\mathrm{im}_h(V_0 \backslash X) \subseteq Y$. However the fact that $D_h \geq 1$ implies that $|\mathrm{im}_h(V_0 \setminus X)| \geq |V_0 \setminus X|$, and hence

$$|S| \;=\; |X| + |Y| \;\geq\; |X| + |\mathrm{im}_h(V_0 \setminus X)| \;\geq\; |V_0|,$$

which is contrary to our assumption about m.

Suppose then that $S \cap V_k \neq \emptyset$, where $1 \leq k \leq h-1$, and write $S \cap V_k = \{s_1, \dots, s_q\}$. Let the remaining elements of S be $s_{q+1}, \dots, s_m$. Take (as allowed by Menger's Theorem) m vertex disjoint paths $\pi_1, \dots \pi_m$ from V_0 to V_h, labelled so that $s_i \in \pi_i$. Observe that $s_j \notin \pi_i$ when $i \neq j$. Let r_i (resp t_i) be the predecessor (resp. successor) of s_i along π_i, so that $r_i \in V_{k-1}$, $t_i \in V_{k+1}$ and r_i, s_i, t_i are consecutive vertices on π_i. Now our assumption that S is bottom heavy means that $\{r_1, \dots, r_q, s_{q+1}, \dots, s_m\}$ is *not* a separating set, and hence there is a path π^* from V_0 to V_h which does not meet this set. π^* meets V_{k-1} in some $r^* \notin \{r_1, \dots, r_q\}$ and, since S is a separating set, π^* meets one of $s_1, \dots, s_q$. Without loss of generality assume that π^* meets s_1, so that $(r^*, s_1) \in E(G)$. Observe that π^* does not meet any other s_i.

Our next claim is that every path in G from $\{r^*, r_1, \dots, r_q\}$ to $\{t_1, \dots, t_q\}$ passes through some s_i, $1 \leq i \leq q$. For paths from some r_i to t_j this is quite clear; if such a path passes through $s^* \notin \{r_1, \dots, r_q\}$ then we can create a path which misses S by following the segment of π_i up to r_i, going to s^* and t_j and then following π_j until it meets V_h. In fact the same argument works for paths from r^* to t_j: here, we follow π^* to r^*, then go through s^* and t_j before following π_j until it meets V_h. The fact that $s_i \notin \pi^*$ for $i \neq 1$ guarantees that this new path misses S. The claim is proved.

We are very near the end of the proof of Proposition 7, but we have not yet used the fact that G is Plünnecke. The previous claim implies that the induced graph G' on vertex set $\{r^*, r_1, \dots, r_q, s_1, \dots, s_q, t_1, \dots, t_q\}$ is Plünnecke; in checking the required properties one uses the Plünneckarity of G to get lots of paths, and these are automatically paths in G' by the claim.

If $v \in V_i$ is a vertex in a Plünnecke graph then write $d^+(v)$ for the number of edges from v into V_{i+1}, and $d^-(v)$ for the number of edges into V_{i-1}. An immediate consequence of the definition of being Plünnecke is that if $(u, v) \in E(G)$ and $v \notin V_h$ then $d^+(u) \geq d^+(v)$, and if $u \notin V_0$ then $d^-(u) \leq d^-(v)$. Look at our Plünnecke graph G'. For any i both (r_i, s_i) and (s_i, t_i) are edges, and hence

$$\sum_i d^+(r_i) \;\geq\; \sum_i d^+(s_i)$$

and

$$\sum_i d^-(s_i) \;\leq\; \sum_i d^-(t_i).$$

However it is clear that

$$d^+(r^*) + \sum_i d^+(r_i) \;=\; \sum_i d^-(s_i)$$

and

$$\sum_i d^+(s_i) \;=\; \sum_i d^-(t_i).$$

The only way to reconcile these two sets of inequalities is to conclude that $d^+(r^*) = 0$, but this is nonsense because we know that (r^*, s_1) is an edge of G'. This final contradiction concludes the proof of Proposition 7. □

The deduction of Theorem 3 from Proposition 4 uses a technique which I call the "taking high powers trick". The best way to illustrate this is to give the derivation. Another example of a similar technique occurs in the paper [10] of Katz and Tao on the Kakeya Problem.

If G, G' are two Plünnecke Graphs of the same level h, with $V(G) = V_0 \cup \dots V_h$ and $V(G') = V_0' \cup \cdots \cup V_h'$ then we define the product graph $G \times G'$ to be the graph on vertex set $\bigcup_{0 \leq i \leq h} V_i \times V_i'$ in which (u, u') is joined to (v, v') by an edge if and only if $(u, v) \in E(G)$, $(u', v') \in E(G')$. It turns out that $G \times G'$ is also Plünnecke; this is a simple exercise which we leave to the reader. A little less obvious is the following.

Proposition 8 *Let G and G' be Plünnecke graphs of level h. Then for any $0 \leq i \leq h$ we have $D_i(G \times G') = D_i(G) \times D_i(G')$.*

Proof. This proposition has the property that one knows that one will have to show that $D_i(G \times G')$ is both no greater and no less than $D_i(G) \times D_i(G')$, and that one such inequality will be trivial. It takes a little thought to realise exactly which one that is, but in fact it is the former. Suppose, then, that $Z \subseteq V_0$ and $Z' \subseteq V_0'$ are such that

$$D_i(G) \;=\; \frac{|\mathrm{im}_i(Z)|}{|Z|}$$

and

$$D_i(G') \;=\; \frac{|\mathrm{im}_i(Z')|}{|Z'|}.$$

The observation that $\mathrm{im}_i(Z \times Z') \subseteq \mathrm{im}_i(Z) \times \mathrm{im}_i(Z')$ then gives

$$D_i(G \times G') \;\leq\; \frac{|\mathrm{im}_i(Z \times Z')|}{|Z \times Z'|} \;\leq\; D_i(G) D_i(G').$$

The other direction requires just a little more thought. Let $X \subseteq V_0 \times V_0'$, and write $X = \bigcup_a(\{a\} \times X_a)$, where the union is over all a for which $X_a \neq \emptyset$ (that is, all a for which there is at least one b with $(a,b) \in X$). If $(a,b) \in X$ and if there is a path in G' from b to d, where $d \in V_i'$, we say that $(a,d) \in X_1$. Observe that

$$|X_1| \;=\; \left|\bigcup_a(\{a\} \times \mathrm{im}_i(X_a))\right| \;=\; \sum_a |\mathrm{im}_i(X_a)| \;\geq\; D_i(G')|X|.$$

Write $X_1 = \bigcup_d(Y_d \times \{d\})$, where again the union is over all d for which $Y_d \neq \emptyset$. Observe that $\mathrm{im}_i(X) = \bigcup_d(\mathrm{im}_i(Y_d) \times \{d\})$, and hence

$$|\mathrm{im}_i(X)| \;=\; \sum_d |\mathrm{im}_d(Y_d)| \;\geq\; D_i(G)\sum_d |Y_d| \;=\; D_i(G)|X_1|.$$

Combining these two inequalities gives $|\mathrm{im}_i(X)| \geq D_i(G)D_i(G')|X|$, and hence we have indeed that $D_i(G \times G') \geq D_i(G) \times D_i(G')$. □

Now let B be a set of size n for which all h-fold sums $b_1 + \cdots + b_h$, $b_i \in B$, are distinct. For example, we could take $B = \{1, 2h, (2h)^2, \ldots, (2h)^{n-1}\}$. We call the graph Plün$(\{0\}, B)$ an *independence graph* of level h and size n, and denote it by $I_h(n)$. All such graphs are isomorphic, and we have $D_i(I_h(n)) = \binom{n+i-1}{i}$ for $1 \leq i \leq h$, this being the number of solutions to the inequalities $1 \leq x_1 \leq x_2 \leq \cdots \leq x_i \leq n$. Observe that we have

$$\frac{n^i}{i!} \;\leq\; D_i(I_h(n)) \;\leq\; n^i \tag{1}$$

To prove Proposition 4 it clearly suffices to show that if G is a Plünnecke graph of level h then $D_i \geq D_h^{i/h}$. This is trivial if $D_h = 0$, and it follows immediately from Proposition 7 if $D_h = 1$. We divide the remaining possibilities into the two cases $0 < D_h < 1$ and $1 < D_h$.

Case 1. $0 < D_h < 1$. Let r be a large integer and let n be the least positive integer such that

$$D_h\left(G^r \times I_h(n)\right) \;\geq\; 1.$$

Using Proposition 8 and (1) we see that $n \leq (h!D_h^{-r})^{1/h} + 1$. Now Proposition 7 implies that $D_i(G^r \times I_h(n)) \geq 1$ which, by another application of Proposition 8 and (1), tells us that

$$D_i^r n^i \;\geq\; 1.$$

It follows that

$$D_i \;\geq\; \left((h!D_h^{-r})^{1/h} + 1\right)^{-i/r},$$

and letting $r \to \infty$ gives the desired inequality.

Case 2. $1 < D_h$. This is very similar to Case 1, except that we use the *inverse independence graph* $I_h(n)^{-1}$. This is the graph obtained by simply reversing $I_h(n)$, and it is obviously Plünnecke by the symmetry of the definition. It is clear that

$$D_h(I_h(n)^{-1}) \; = \; \binom{n+h-1}{h}^{-1} \; \geq \; n^{-h}, \tag{2}$$

and that we have the upper bound

$$D_i(I_h(n)^{-1}) \; \leq \; \frac{\binom{n+h-i-1}{h-i}}{\binom{n+h-1}{h}} \; \leq \; h! n^{-i}. \tag{3}$$

Given a positive integer r let n be maximal so that

$$D_h(G^r \times I_h(n)^{-1}) \; \geq \; 1.$$

Using Proposition 8 and (2) we see that $n \geq D_h^{r/h} - 1$. Now Proposition 7 implies that $D_i(G^r \times I_h(n)^{-1}) \geq 1$ which, by another application of Proposition 8 and (1), tells us that

$$D_i^r \; \geq \; n^i/h!.$$

It follows that

$$D_i \; \geq \; \frac{\left(D_h^{r/h} - 1\right)^{i/r}}{(h!)^{1/r}},$$

and letting $r \to \infty$ once again gives the desired inequality. This at last concludes the proof of Proposition 4 and hence, by our earlier work, of Theorem 3. □

1.3 Application: Freiman's theorem for groups with bounded torsion.

Let G be a group in which every element has order at most r, and let $A \subseteq G$ have small sumset. In this section we prove an analogue of Freiman's theorem in this setting due to Ruzsa [13]. It turns out that the bounded torsion enables us to give a proof which is much simpler than the proof of Freiman's theorem proper (which will be the subject of our second and third lectures).

Theorem 9 (Freiman in torsion groups) *Let G be a group in which every element has order at most r, and let $A \subseteq G$ be a set with $|A + A| \leq C|A|$. Then A is contained within a coset of some subgroup H of G with $|H| \leq C^2 r^{C^4}|A|$.*

Proof. Let $|A| = n$. The important trick comes straight away. Let $X \subseteq 2A - A$ be maximal subject to the condition that the sets $A + x$, $x \in X$, are disjoint. Observe that all the sets $A+x$ lie in $3A-A$ which has size at most $C^4 n$ by Plünnecke. It follows that $|X| \leq C^4$. Now if $t \in 2A - A$ then $A + t$ intersects $A + x$ for some $x \in X$; indeed if $t \in X$ this is trivial, and if $t \notin X$ then $A+t$ is certainly not disjoint from $\bigcup_{x \in X}(A+x)$ by maximality. It follows that $t \in X + (A - A)$, that is

$$2A - A \ \subseteq \ X + (A - A).$$

Adding A to both sides of this inclusion gives

$$3A - A \ \subseteq \ X + (2A - A) \ \subseteq \ 2X + (A - A),$$

and similarly (or by induction)

$$iA - A \ \subseteq \ (i-1)X + (A - A)$$

for any $i \geq 2$. Let H be the subgroup of G generated by A. It is clear that

$$H \ = \ \bigcup_{i \geq 1}(iA - A),$$

and hence H is contained in $I + (A - A)$ where I is the subgroup generated by X. The fact that G has bounded torsion implies that everything in I can be written as

$$n_1 x_1 + \cdots + n_{|X|} x_{|X|},$$

where the x_j are the elements of X and $0 \leq n_j < r$. It follows that $|I| \leq r^{|X|} \leq r^{C^4}$. Finally we have

$$|H| \ \leq \ |I||A - A| \ \leq \ C^2 n r^{C^4},$$

which completes the proof. Note that the trick used here, namely a clever choice of something "maximal", is very simple yet extremely powerful. We will see it again in Lecture 2. □

2 Lecture 2: Fourier analysis, Chang's structure theorem and the structure of sumsets

In this lecture we will use the discrete Fourier transform to investigate the structure of $2A-2A$, where $A \subseteq \mathbb{Z}_N$ has small sumset (throughout the lecture, N will be an odd prime). This will require us to use a wide range of ideas, the most interesting of which is the concept of *dissociativity.* The main result of this lecture is a key step in the proof of Freiman's theorem. To state it we need a definition.

Definition 10 *Let $R \subseteq \mathbb{Z}_N$, and let $\delta > 0$ be a positive real number. Then the* Bohr Neighbourhood $B(R,\delta)$ *is defined to be the set*

$$\left\{x \;\middle|\; \left\|\frac{rx}{N}\right\| \leq \delta \textit{ for all } r \in R\right\}.$$

The notation $\|x\|$ refers to the distance from x to the nearest integer.

The next theorem is the main result of this lecture, and is the only result that we shall carry through to the third lecture.

Theorem 11 *Let $A \subseteq \mathbb{Z}_N$ have cardinality αN and suppose that $|A+A| \leq C|A|$. Then $2A-2A$ contains some Bohr neighbourhood $B(K,\delta)$, where $|K| \leq 8C\log(1/\alpha)$ and $\delta \geq (160C\log(1/\alpha))^{-1}$.*

To set this in context we remark that in applications α will be a power of C^{-1}, so that K and δ^{-1} are both of the order $C\log C$.

2.1 The discrete Fourier transform.

Let $f,g : \mathbb{Z}_N \to \mathbb{R}$ be any functions. Write $\omega = e^{2\pi i/N}$ and define the Fourier transform of f by

$$\hat{f}(r) \;=\; \sum_x f(x)\omega^{rx}$$

for any $r \in \mathbb{Z}_N$. It is an easy matter to check the following proposition.

Proposition 12 (Fourier formulæ) *Let $f,g : \mathbb{Z}_N \to \mathbb{R}$. Then we have*

(i) (Inversion) *$f(x) = N^{-1}\sum_r \hat{f}(r)\omega^{-rx}$;*

(ii) (Parseval) *$\sum_x f(x)g(x) = N^{-1}\sum_r \hat{f}(r)\overline{\hat{g}(r)}$;*

(iii) (Convolution) *Define $f*g(x) = \sum_y f(y)g(y-x)$. Then $(f*g)\hat{}(r) = \hat{f}(r)\overline{\hat{g}(r)}$.*

We will frequently take the liberty of identifying sets with their characteristic functions. Hence if $A \subseteq \mathbb{Z}_N$ then $A(x) = 1$ or 0 according as $x \in A$ or not. To get used to the notation and the above proposition, we note some expressions that will be required later. First of all observe that an immediate consequence of Parseval's identity is the formula

$$N^{-1} \sum_r |\hat{A}(r)|^2 \; = \; |A|. \tag{4}$$

Secondly, by (ii) and (iii) of the proposition we have

$$N^{-1} \sum_r |\hat{A}(r)|^4 \; = \; \# \left\{ (a_1, a_2, a_3, a_4) \in A^4 \;\middle|\; a_1 + a_2 = a_3 + a_4 \right\}. \tag{5}$$

We call this the number of *additive quadruples* of A. Thirdly note that $A * A * A * A(x)$ is the number of quadruples $(a_1, a_2, a_3, a_4) \in A^4$ with $a_1 + a_2 - a_3 - a_4 = x$, and so $A * A * A * A(x) > 0$ if and only if $x \in 2A - 2A$. Using (i) and (iii) of the proposition we have that $x \in 2A - 2A$ if and only if

$$\sum_r |\hat{A}(r)|^4 \omega^{-rx} \; > \; 0. \tag{6}$$

2.2 Dissociativity and the Fourier transform.

It is quite likely that I will not have had time to cover the topics of this section fully in the lecture. The objective is to prove the following theorem of Chang [3]:

Theorem 13 *Let $\rho, \alpha \in [0, 1]$, let $A \subseteq \mathbb{Z}_N$ be a set of size αN and let $R \subseteq \mathbb{Z}_N$ be the set of all r for which $|\hat{A}(r)| \geq \rho |A|$. Then R is contained in a cube of dimension at most $2\rho^{-2} \log(1/\alpha)$.*

(If $\Lambda = \{\lambda_1, \dots, \lambda_k\}$ is a subset of an abelian group then write $\overline{\Lambda}$ for the set of everything of the form $\sum_j \epsilon_j \lambda_j$ where $\epsilon_j \in \{-1, 0, 1\}$. We call $\overline{\Lambda}$ the *cube* spanned by Λ, and deem k to be its dimension.)

The point of Theorem 13 is that the "large spectrum" of A, that is the set of points at which $\hat{A}$ is large, is very highly structured. Observe that Parseval's identity gives only the inequality $|R| \leq \rho^{-2} \alpha^{-1}$, which is much weaker than Chang's theorem for small α.

We say that a set $\Lambda = (\lambda_j)_{j=1}^k \subseteq \mathbb{Z}_N$ is *dissociated* if the only solution to

$$\sum_j \epsilon_j \lambda_j \; = \; 0$$

with $\epsilon_j \in \{-1, 0, 1\}$ is the trivial one in which $\epsilon_j = 0$ for all j.

Suppose in what follows that Λ is a dissociated set. We will consider cosine polynomials of the form

$$f(x) \;=\; \sum_{j=1}^{k} c_j \cos\left(\frac{2\pi\lambda_j x}{N} + \beta_j\right) \tag{7}$$

with frequencies in Λ. Here $c_j \in \mathbb{R}$ and $\beta_j \in \mathbb{T}$ for $j = 1, \dots, k$. Our objective is to show that such polynomials behave rather like sums of independent random variables, and to derive a combinatorial consequence of this fact. We shall establish this behaviour by remodelling a classical technique of Bernstein which is nearly 80 years old. First of all we isolate three simple lemmas from the proof.

Lemma 14 *Let t be any real number and suppose that y is a real number with $|y| \leq 1$. Then we have*

$$e^{ty} \;\leq\; \cosh(t) + y\sinh(t).$$

Proof. Observe that the function $g(x) = e^{tx}$ is convex on $[-1, 1]$, having non-negative derivative. It follows that

$$g(y) \;\leq\; \left(\frac{1-y}{2}\right) g(-1) + \left(\frac{1+y}{2}\right) g(1).$$

The result follows immediately. □

Lemma 15 *Let u be any real number. Then we have the inequality*

$$\cosh(u) \;\leq\; e^{u^2/2}.$$

Proof. This follows by comparing the powers series of the two sides term by term. □

The following is nothing more than Parseval's identity for cosine polynomials.

Lemma 16 *Let f be a cosine polynomial of the form* (7)*. Then we have $\sum f(x)^2 = \frac{N}{2}\sum_j c_j^2$.*

Proof. By writing each cosine $\cos\left(\frac{2\pi\lambda_j x}{N} + \beta_j\right)$ as a linear combination $\gamma_j e(\lambda_j x/N) + \overline{\gamma}_j e(-\lambda_j x/N)$ where $|\gamma_j| = \frac{1}{2}$ one can check that

$$\sum_x \cos\left(\frac{2\pi\lambda_i x}{N} + \beta_i\right) \cos\left(\frac{2\pi\lambda_j x}{N} + \beta_j\right) \;=\; 0$$

unless $i = j$, in which case it equals $N/2$. Here we have used dissociativity in a very minor way in observing that $\lambda_i + \lambda_j \neq 0$. The result follows immediately. □

Proposition 17 *Let $t \in \mathbb{R}$ and let f be a cosine polynomial of the form* (7)*. Then we have the inequality*

$$N^{-1}\sum_x \exp(tf(x)) \;\le\; \exp\left(N^{-1}t^2\sum_x f(x)^2\right).$$

Proof. Lemma 14 implies that

$$N^{-1}\sum_x \exp(tf(x)) \;\le\; N^{-1}\sum_x\prod_{j=1}^k\left(\cosh(tc_j)+\sinh(tc_j)\cos\left(\frac{2\pi\lambda_j x}{N}+\beta_j\right)\right). \qquad (8)$$

As in the proof of Lemma 16 write each cosine $\cos\left(\frac{2\pi\lambda_j x}{N}+\beta_j\right)$ as a linear combination $\gamma_j e(\lambda_j x/N)+\overline{\gamma}_j e(-\lambda_j x/N)$. If one does this in the right- hand side of (8) and multiplies out one gets a huge linear combination (with coefficients depending on t) of exponentials

$$e\left(2\pi(\epsilon_1\lambda_1+\cdots+\epsilon_k\lambda_k)x/N\right),$$

where each ϵ_j is $-1, 0$ or 1. The dissociativity of Λ implies that the integral over $\mathbb{Z}_N$ of all but one of these terms vanishes. The remaining term is the one with $\epsilon_1=\cdots=\epsilon_k=0$, which comes with a coefficient $\prod_{j=1}^k\cosh(tc_j)$. Thus from (8), Lemma 15 and Lemma 16 we deduce that

$$\begin{aligned} N^{-1}\sum_x \exp(tf(x)) &\le \prod_{j=1}^k \cosh(tc_j) \\ &\le \exp\left(\frac{1}{2}t^2\sum_j c_j^2\right) \\ &= \exp\left(N^{-1}t^2\sum f(x)^2\right). \end{aligned}$$

as required. □

Proposition 18 (Chang) *Let $\rho, \alpha \in [0,1]$, let $A \subseteq \mathbb{Z}_N$ be a set of size αN and let $R \subseteq \mathbb{Z}_N$ be the set of all r for which $|\hat{A}(r)| \ge \rho|A|$. Let Λ be a dissociated subset of R. Then $|\Lambda| \le 2\rho^{-2}\log\left(\frac{1}{\alpha}\right)$.*

Proof. Suppose that $\Lambda = \{\lambda_j\}_{j=1}^k$ and set

$$f(x) \;=\; \Re\left(\sum_{j=1}^k \hat{A}(\lambda_j)\omega^{-\lambda_j x}\right).$$

This is a cosine polynomial of the type described by Proposition 17 in which $c_j = |\hat{A}(\lambda_j)|$. Observe also that $\hat{f}(r) = N\hat{A}(r)/2$ if $r \in \Lambda \cup -\Lambda$ and zero otherwise. This implies that

$$\sum_x f(x)A(x) \;=\; N^{-1}\sum_r \hat{f}(r)\overline{\hat{A}(r)} \;=\; 2N^{-2}\sum_r |\hat{f}(r)|^2 \;=\; 2N^{-1}\sum_x |f(x)|^2. \tag{9}$$

Proposition 17, an application of the weighted AM-GM inequality and (9) give

$$\begin{aligned}\frac{1}{|A|}\exp\left(\frac{t^2\sum|f(x)|^2}{N}\right) \;&\geq\; \frac{1}{N|A|}\sum_x \exp(tf(x)) \\ &\geq\; \frac{1}{N|A|}\sum \exp(tf(x))A(x) \\ &\geq\; \frac{1}{N}\exp\left(\frac{t}{|A|}\sum f(x)A(x)\right). \\ &=\; \frac{1}{N}\exp\left(\frac{2t}{n|A|}\sum_x f(x)^2\right).\end{aligned}$$

Choosing $t = |A|^{-1}$ gives

$$\alpha \;\geq\; \exp\left(\frac{1}{N|A|^2}\sum_x f(x)^2\right).$$

The proof of the proposition is concluded by using Lemma 16 and our assumption that $\Lambda \subseteq R$ to observe that

$$\begin{aligned}\sum_x f(x)^2 \;&=\; \frac{N}{2}\sum_j c_j^2 \\ &=\; \frac{N}{2}\sum_{r\in\Lambda}|\hat{A}(r)|^2 \\ &\geq\; \frac{k\rho^2 N|A|^2}{2}.\end{aligned}$$

Theorem 13 is a simple corollary of this proposition. Indeed let A, α, ρ be as in the statement of the theorem and let Λ be a *maximal* dissociated subset of R. Then $|\Lambda| \leq 2\rho^{-2}\log(1/\alpha)$. The maximality of Λ implies that any $r \in R$ is involved in some equation

$$\epsilon r + \sum_i \epsilon_i\lambda_i \;=\; 0;$$

otherwise, we could add r to Λ to create a larger dissociated subset of R. Thus R is contained in the cube spanned by Λ. □

It turns out that the exponential moment inequality of Proposition 17 implies good bounds for $\|f\|_p$. These bounds can also be established directly, and may then be used to deduce Theorem 13 in another way. See [3] or [8] for details.

Lemma 19 *Let f be as in* (7) *and let $p \geq 2$. Then we have $\|f\|_p \leq 5\sqrt{p}\|f\|_2$.*

Proof. Suppose that $t \geq 0$. Suppose to begin with that p is an even integer. A simple calculus exercise confirms that $y \mapsto y^p e^{-ty}$ is maximised on $[0, \infty)$ when $y = p/t$ and it follows that one has

$$y^p \;\leq\; e^{ty}\left(\frac{p}{t}\right)^p.$$

Substituting $y = f(x)$ and using Proposition 17 gives

$$\|f\|_p^p \;\leq\; \left(\frac{p}{t}\right)^p e^{t^2\|f\|_2^2}.$$

Putting $t = \sqrt{\frac{p}{2\|f\|_2^2}}$ gives the inequality $\|f\|_p \leq 3\sqrt{p}\|f\|_2$. Obtaining an inequality for any real $p \geq 2$ is now an easy matter. Indeed given p there is an even integer $p' \leq 2p$, and then

$$\|f\|_p \;\leq\; \|f\|_{p'} \;\leq\; 5\sqrt{p}\|f\|_2.$$

2.3 Proof of Theorem 11.

Let $A \subseteq \mathbb{Z}_N$ be a set of size αN with $|A + A| \leq C|A|$. Our first observation is that A has many additive quadruples, and hence by (5) that $\sum_r |\hat{A}(r)|^4$ is large. To see this write $r_A(x)$ for the number of pairs $(a_1, a_2) \in A^2$ with $a_1 + a_2 = x$. Then

$$\begin{aligned} N^{-1}\sum_r |\hat{A}(r)|^4 &= \#\text{ additive quadruples in } A \\ &= \sum_{x \in A+A} r_A(x)^2 \\ &\geq \frac{1}{|A+A|}\left(\sum_x r_A(x)\right)^2 \\ &= \frac{|A|^4}{|A+A|} \\ &= \frac{|A|^3}{C}. \end{aligned}$$

Thus

$$\sum_r |\hat{A}(r)|^4 \;\geq\; \frac{\alpha^3 N^4}{C}. \tag{10}$$

Now let R be the set of all $r \neq 0$ for which $|\hat{A}(r)| \geq |A|/2\sqrt{C}$. We claim that $2A - 2A$ contains $B(R, \frac{1}{20})$. Indeed if $x \in B(R, \frac{1}{20})$ then, for any $r \in R$,

$$\left|1 - \omega^{-rx}\right| \;=\; \left|1 - e^{-2\pi irx/N}\right| \;=\; 2\left|\sin(\pi rx/N)\right| \;\leq\; \frac{2\pi}{20} \;<\; \frac{1}{2}.$$

It follows that

$$\begin{aligned}\sum_r |\hat{A}(r)|^4\omega^{-rx} &= \sum_r |\hat{A}(r)|^4 - \sum_r |\hat{A}(r)|^4\left(1 - \omega^{-rx}\right)\\ &> \frac{1}{2}\sum_r |\hat{A}(r)|^4 - 2\sum_{r\notin R, r\neq 0} |\hat{A}(r)|^4\\ &\geq \frac{\alpha^3N^4}{2C} - 2\sup_{r\notin R, r\neq 0} |\hat{A}(r)|^2 \sum_r |\hat{A}(r)|^2\\ &\geq 0.\end{aligned}$$

Here we have made use of (4) and (10). The claim follows immediately from this and (6). Parseval's identity tells us that $|R| \leq 4C/\alpha$. However we did not prove Theorem 13 for nothing, and using it allows us to get something much stronger. The theorem implies that $R \subseteq \overline{\Lambda}$, where $|\Lambda| \leq 8C\log(1/\alpha)$. This means that $B(R, \frac{1}{20})$ contains $B\left(\Lambda, \frac{1}{20|\Lambda|}\right)$; indeed any $r \in R$ can be written as $\sum_{j=1}^{|\Lambda|} \epsilon_j\lambda_j$ with $\epsilon_j \in \{-1, 0, 1\}$, and so if $\|\lambda_j x/N\| \leq 1/20|\Lambda|$ for all x we have

$$\left\|\frac{rx}{N}\right\| \;\leq\; \sum_{j=1}^{|\Lambda|}\left\|\frac{\lambda_j x}{N}\right\| \;\leq\; \frac{1}{20}.$$

Theorem 11 follows from these remarks and the claim immediately preceding them. □

2.4 Further remarks

It is natural to ask whether Theorem 13 is best possible, in the sense of whether or not one can significantly add to the structural information it gives about the large spectrum. At least, I thought this was a natural question and that is why I bothered to prove the following:

Theorem 20 (G. [7]) *Let α, ρ be positive real numbers satisfying $\alpha \leq 1/8$, $\rho \leq 1/32$ and*

$$\rho^{-2}\log(1/\alpha) \;\leq\; \frac{\log N}{\log\log N}. \tag{11}$$

Then there is a set $A \subseteq \mathbb{Z}_N$ with $|A| = \lfloor \alpha N\rfloor$ such that $|\hat{A}(r)| \geq \rho|A|$ for all $r \in R$, where R is not contained in $\overline{\Lambda}$ for any set Λ with $|\Lambda| \leq 2^{-12}\rho^{-2}\log(1/\alpha)$.

3 Lecture 3: Minkowski's second theorem. Freiman homomorphisms. Conclusion of the proof.

The main result of the previous lecture was Theorem 11, which said that if $A \subseteq \mathbb{Z}_N$ has small sumset then $2A - 2A$ contains a large Bohr neighbourhood. To use this to prove Freiman's theorem we must understand three further issues:

- What is the structure of a Bohr neighbourhood? In particular, in what sense does it look like a multidimensional progression?
- Freiman's theorem concerns subsets of $\mathbb{Z}$, not of $\mathbb{Z}_N$. How can we move between the two situations?
- What on earth has the structure of $2A - 2A$ got to do with the structure of A?

We will answer these questions in this final lecture, and this will conclude the proof of Freiman's theorem.

3.1 Minkowski's second theorem and the structure of Bohr neighbourhoods

It is clear that a Bohr neighbourhood $B(\{r\}, \delta)$ defined by one point is simply an arithmetic progression in $\mathbb{Z}_N$. The structure of higher- dimensional Bohr neighbourhoods is rather less obvious, but as it turns out they resemble high-dimensional arithmetic progressions. In this section we establish the following quantitative result along these lines using a result from the geometry of numbers.

Proposition 21 *Let $R \subseteq \mathbb{Z}_N$ be a set of cardinality k, and let $\delta \in (0, 1/2)$. Then the Bohr neighbourhood $B(R, \delta)$ contains a proper arithmetic progression of dimension k and size at least $(\delta/k)^k N$.*

The fact that the progression we obtain is proper turns out to be very helpful later on.

Let $\Lambda \subseteq \mathbb{R}^n$ be a lattice, that is a set of the form $\bigoplus_{j=1}^n \mathbb{Z}v_j$ where the v_j are linearly independent. If $K \subseteq \mathbb{R}^n$ is an open convex body then we define the kth *successive minimum* of K with respect to Λ to be

$$\lambda_k(K, \Lambda) \; = \; \inf\{\lambda > 0 \mid \lambda K \text{ contains } k \text{ linearly independent elements of } \Lambda\}.$$

Minkowski's second theorem is the following rather remarkable result.

Theorem 22 (Minkowski's second theorem) *Suppose that K is an open, convex, centrally symmetric subset of $\mathbb{R}^n$ and let Λ be a lattice. Let $\lambda_i = \lambda_i(K, \Lambda)$ be the successive minima of K with respect to Λ. Then we have*

$$\lambda_1\lambda_2 \ldots \lambda_n|K| \; \leq \; 2^n|\Lambda|.$$

Here $|K|$ is the volume of K and $|\Lambda|$ is the determinant of Λ, that is the volume of the parallelepipid spanned by $v_1, \dots, v_n$ (or equivalently $|\det(v_1, \dots, v_n)|$).

Minkowski's second theorem easily implies his first, which says that if K is a closed, convex, centrally symmetric body and if $|K| \geq 2^n|\Lambda|$ then K contains a non-zero element of Λ.

Now given an open convex body K and a lattice Λ we may use the successive minima to pick a basis $b_1, \dots, b_n$ for $\mathbb{R}^n$. We build this basis inductively; first of all pick an element of $\lambda_1\overline{K} \cap \Lambda$ and call it b_1, then pick an element of $\lambda_2\overline{K} \cap \Lambda$ which is not in the linear span of b_1, call it b_2, and so on. I call such a basis a *directional basis* with respect to the pair (K, Λ). (this is not standard terminology). A very convenient consequence of our assumption that K is open is that $b_j \notin \lambda_j K$, so that every lattice point in $\lambda_j K$ is a linear combination of $b_1, \dots, b_{j-1}$.

There now follows a proof of Minkowski's second theorem. I have to confess to not really understanding this proof on a conceptual level, and would welcome any insights that anyone might have about just where the argument comes from. We begin with a lemma of Blichfeldt.

Lemma 23 (Blichfeldt) *Let K be any open body with $|K| > |\Lambda|$. Then K contains two distinct points a and b with $a - b \in \Lambda$.*

Proof. This is, at heart, an extremely simple averaging argument. We may assume that K is bounded, since $|K \cap B(0, R_1)| > |\Lambda|$ for R_1 sufficiently large. Suppose that the lemma is false, so that any translate $K + x$ contains at most one point of Λ. Let C be the ℓ^∞ cube $\{x \mid \|x\| \leq R_2\}$. Then certainly we have

$$\frac{1}{|C|}\int_{x \in C} |(K + x) \cap \Lambda| \, dx \;\leq\; 1, \tag{12}$$

and furthermore

$$\frac{1}{|C|}\int_{x \in C} |(K + x) \cap \Lambda| \, dx \;=\; \int K(y) \left(\frac{1}{|C|}\int_{x \in C} \Lambda(x - y) \, dx \right) dy. \tag{13}$$

However it is "obvious" (i.e. it is left as an exercise to the reader) that

$$\lim_{R_2 \to \infty} \frac{1}{|C|}\int_{x \in C} \Lambda(x - y) \, dy \;=\; \frac{1}{|\Lambda|}$$

uniformly for y in any compact set. Since, in particular, K is contained in a compact set we may take limits of (12) and (13) as $R_2 \to \infty$ to derive a contradiction. □

Proceeding now with the proof of Theorem 22, let K be an open, centrally symmetric and convex subset of $\mathbb{R}^n$ and let Λ be a fixed lattice. Let $\lambda_1, \dots, \lambda_n$ be the successive minima of K with respect to Λ and let $b_1, \dots, b_n$ be the corresponding directional basis. For each $j \in \{1, \dots, n\}$ we define a map $\phi_j : K \to K$ by mapping $x \in K$ to the centre of gravity of the slice of K which contains x and is parallel to the subspace spanned by $b_1, \dots, b_{j-1}$ (for $j = 1$, $\phi_1(x) = x$). Define a map $\phi : K \to \mathbb{R}^n$ by

$$\phi(x) \;=\; \sum_{j=1}^{n} (\lambda_j - \lambda_{j-1})\phi_j(x), \tag{14}$$

where we are operating with the convention that $\lambda_0 = 0$.

Let us make a few further observations concerning the ϕ_j and ϕ. We define functions $c_{ij} : \mathbb{R}^n \to \mathbb{R}$ by looking at ϕ_j in coordinates relative to the directional basis $b_1, \dots, b_n$. Specifically if $x = x_1 b_1 + \dots + x_n b_n$ we write

$$\phi_j(x) \;=\; \sum_i c_{ij}(x) b_i. \tag{15}$$

The definition of ϕ_j means that $c_{ij}(x) = x_i$ whenever $i \geq j$, whilst if $j > i$ then $c_{ij}(x)$ depends only on $x_{i+1}, \dots, x_n$. It follows that we can write

$$\phi(x) \;=\; \sum_{i=1}^{n} b_i \left(\lambda_i x_i + \psi\left(x_{i+1}, \dots, x_n\right)\right) \tag{16}$$

for certain continuous functions ψ_j.

Now observe that $|\phi(K)| = \lambda_1 \dots \lambda_n |K|$, the determinant of the Jacobean of the transformation $x_i' = \lambda_i x_i + \psi_i(x_{i+1}, \dots, x_n)$ being particularly easy to evaluate due to the matrix being upper triangular. Suppose, as a hypothesis for contradiction, that $\lambda_1 \dots \lambda_n |K| > 2^n |\Lambda|$. By Blichfeldt's lemma and the preceding observation this means that $\phi(K)$ contains two elements $\phi(x)$ and $\phi(y)$ which differ by an element of 2Λ, so that $\frac{1}{2}(\phi(x) - \phi(y)) \in \Lambda$. Write $x = \sum x_i b_i$ and $y = \sum y_i b_i$, and suppose that k is the largest index such that $x_k \neq y_k$. Then we have $\phi_i(x) = \phi_i(y)$ for $i > k$, so that

$$\begin{aligned} \frac{\phi(x) - \phi(y)}{2} \;&=\; \sum_{j=1}^{n} (\lambda_j - \lambda_{j-1}) \left(\frac{\phi_j(x) - \phi_j(y)}{2}\right) \\ &=\; \sum_{j=1}^{k} (\lambda_j - \lambda_{j-1}) \left(\frac{\phi_j(x) - \phi_j(y)}{2}\right). \end{aligned}$$

This has two consequences. First of all the convexity of K implies that $\frac{1}{2}(\phi_j(x) - \phi_j(y)) \in K$ for all j, and hence (again by convexity) $\frac{1}{2}(\phi(x) - \phi(y)) \in \lambda_k K$. Secondly we may easily evaluate the coefficient of b_k when $\frac{1}{2}(\phi(x) - \phi(y))$ is written in terms of our directional basis.

It is exactly $\lambda_k(x_k - y_k)/2$. In particular this is non-zero, which is contrary to our earlier observation that $\Lambda \cap \lambda_k K$ is spanned by $b_1, \dots, b_{k-1}$. $\square$

I have already professed my lack of intuition for this argument. There is, however, one small mystery that can be cleared up. This is our use of the openness of K, which at first sight seems to have been crucial. In fact this is just a convenience, and one could have dealt with closed bodies throughout. Doing that would have required the introduction of various messy devices, such as the consideration of a δ-neighbourhood of K, and the use of openness is far prettier.

Let us now deduce Theorem 21. Let $R = \{r_1, \dots, r_k\}$ be a subset of $\mathbb{Z}_N$ and consider the lattice $\Lambda = N\mathbb{Z}^k + (r_1, \dots, r_k)\mathbb{Z}$. This is a slight abuse of notation, for the r_i are not integers, but the definition should be clear and unambiguous. We have $|\Lambda| = N^{k-1}$. Let K be the ℓ^∞ box $\{x \in \mathbb{R}^n : \|x\|_\infty < \delta N\}$, which is open, convex and has volume $(2\delta)^k N^k$, and let $b_1, \dots, b_k$ be a directional basis for $\mathbb{R}^k$. We have $b_i \in \lambda_i \overline{K}$, where $\lambda_1, \dots, \lambda_k$ are the successive minima of K. Now b_i lies in Λ, and so it has the form

$$b_i \;=\; x_i(r_1, \dots, r_k) + Nv,$$

where $v \in \mathbb{Z}^k$ and, by another abuse of notation, we regard the x_i as elements of $\mathbb{Z}_N$. The fact that $\|b_i\|_\infty \leq \lambda_i \delta N$ implies that each $\|x_i r/N\|$, $r \in R$, is at most $\lambda_i \delta$. Look at the multidimensional AP

$$Q \;=\; \{\mu_1 x_1 + \dots + \mu_k x_k, |\mu_i| \leq \lfloor 1/k\lambda_i \rfloor\} \;\subseteq\; \mathbb{Z}_N.$$

We claim that $Q \subseteq B(R, \delta)$. Indeed for any $r \in R$ we have

$$\begin{aligned} \left\| \frac{r(\mu_1 x_1 + \dots + \mu_k x_k)}{N} \right\| &\leq \sum_{i=1}^k |\mu_i| \left\| \frac{r x_i}{N} \right\| \\ &\leq \sum_{i=1}^k \left\lfloor \frac{1}{\lambda_i k} \right\rfloor \lambda_i \delta \\ &\leq \delta. \end{aligned}$$

The size of Q is at least $k^{-k}(\lambda_1 \dots \lambda_k)^{-1}$, which, by Minkowski's Second Theorem, is at least $(\delta/k)^k N$. It remains to show that Q is proper. Suppose that

$$\mu_1 x_1 + \dots + \mu_k x_k = \mu_1' x_1 + \dots + \mu_k' x_k$$

in $\mathbb{Z}_N$, where $|\mu_i|, |\mu_i'| \leq \lfloor 1/k\lambda_i \rfloor$. Then the vector

$$b \;=\; (\mu_1 - \mu_1')b_1 + \dots + (\mu_k - \mu_k')b_k$$

lies in $N\mathbb{Z}^k$ and furthermore

$$\begin{aligned}\|b\|_\infty &\leq \sum_{i=1}^{k} 2 \left\lfloor \frac{1}{\lambda_i k} \right\rfloor \|b_i\|_\infty \\ &\leq 2\delta N.\end{aligned}$$

Since we are assuming that $\delta < 1/2$ it follows that $b = 0$ and hence, due to the linear independence of the b_i, that $\mu_i = \mu_i'$ for all i. Therefore Q is indeed proper. □

Remark. It is interesting to enquire as to how accurate Proposition 21 actually is. If k is small and $R = \{r_1, \dots, r_k\}$ is selected "randomly" then the events $\|xr_i/N\|$ should be roughly independent. Each has probability about 2δ, so we expect $|B(R,\delta)| \approx (2\delta)^k N$. Thus the dependence on δ in Proposition 21 is certainly best possible. It is likely that the factor of k is necessary as well; I have not though seriously about this point.

3.2 Passing from $\mathbb{Z}$ to $\mathbb{Z}_N$: Freiman homomorphisms and another argument of Ruzsa.

In the book [4] Freiman introduces a number of tools for studying the structure of sets under addition. One that has turned out to be very useful is (what is now known as) the notion of *Freiman homomorphism.* Let k be a positive integer, let A be a subset of an abelian group G, and let $\phi : A \rightarrow H$ be a function from A into another abelian group. We say that ϕ is a (Freiman) k-homomorphism if whenever $x_1, x_2, \dots, x_{2k}$ are elements of A with

$$x_1 + \cdots + x_k = x_{k+1} + \cdots + x_{2k}$$

we have

$$\phi(x_1) + \cdots + \phi(x_k) = \phi(x_{k+1}) + \cdots + \phi(x_{2k}).$$

If ϕ has an inverse which is also a Freiman homomorphism then we say that it is a Freiman isomorphism. We will occasionally use the notation $A \cong_k B$ to express the fact that A is Freiman k-isomorphic to B. Observe that a Freiman homomorphism induces a well-defined function on the k-fold sumset kA in an obvious way. Observe also that a 1-1 Freiman homomorphism need not be a Freiman isomorphism; for example the obvious map from $\{0,1\} \subseteq \mathbb{Z}$ to $\mathbb{Z}_2$ is a 1-1 Freiman homomorphism of all orders, but ϕ^{-1} is not a Freiman homomorphism of any order $k \geq 2$. This slightly disconcerting property would doubtless make an algebraist use a different nomenclature, but we shall stick with the standard practice.

The following proposition provides the link between subsets of $\mathbb{Z}$ and subsets of $\mathbb{Z}_N$ that we are looking for.

Proposition 24 (Ruzsa) *Let $A \subseteq \mathbb{Z}$ be a set of size n with $|A+A| = Cn$. Let $m > C^{2k}n$, and let $k \geq 2$ be an integer. Then there is a subset $A' \subseteq A$ of size at least n/k which is k-isomorphic to a subset of $\mathbb{Z}_m$.*

Proof. Let p be a very large prime number, and consider the composition of maps

$$\mathbb{Z} \xrightarrow{\psi_1} \mathbb{Z}_p \xrightarrow{\psi_2(q)} \mathbb{Z}_p \xrightarrow{\psi_3} \mathbb{Z} \xrightarrow{\psi_4} \mathbb{Z}_m$$

where ψ_1 and ψ_4 are reduction mod p and m respectively, $\psi_2(q)$ is multiplication by q and ψ_3 is the map which sends $x \in \mathbb{Z}_p$ to the corresponding residue in the interval $\{0, 1, \ldots, p-1\}$.

ψ_1, ψ_2 and ψ_4 are Freiman homomorphisms of any order. ψ_3 is a homomorphism of order k when restricted to any interval of the form $I_j = \left(\frac{j-1}{k}p, \frac{j}{k}p\right]$. Choose p large enough that $\psi_1|_A$ is 1-1. Write $S_j = \{x \in A \mid \psi_2(\psi_1(x)) \in I_j\}$. Then, for any q, there is $j = j(q)$ such that $|S_j| \geq n/k$. Observe that for any q the composition $\psi = \psi_1 \circ \psi_2(q) \circ \psi_3 \circ \psi_4$ is a k-homomorphism when restricted to $S_{j(q)}$.

To conclude the proof we show that there is a choice of q for which ψ is invertible, and for which its inverse is also a k-homomorphism. It suffices to show that whenever

$$\psi(x_1) + \cdots + \psi(x_k) \;=\; \psi(x_{k+1}) + \cdots + \psi(x_{2k})$$

we have

$$x_1 + \cdots + x_k \;=\; x_{k+1} + \cdots + x_{2k},$$

since this clearly implies that ψ is 1-1. The only way in which these conditions can fail to hold, for a given q, is if there is some non-zero expression $s = x_1 + x_2 + \cdots + x_k - x_{k+1} - \cdots - x_{2k}$ such that

$$qs(\mathrm{mod} p) \;\equiv\; 0(\mathrm{mod} m), \tag{17}$$

where the $(\mathrm{mod} p)$ instructs one to take the least non-negative residue modulo p. Let us fix s and ask about values of q for which (17) fails. As q ranges over $\mathbb{Z}_p^\times$, $qs(\mathrm{mod} p)$ covers $[1, \ldots, p-1]$. The number of elements in this interval divisible by m is at most $(p-1)/m$. Now each s lies in the set $(kA - kA) \setminus \{0\}$, and by Plünnecke's inequality this has cardinality less than $C^{2k}n$. It follows that the number of "bad" q, that is q such that (17) holds for some s, is less than $C^{2k}n(p-1)/m < p-1$. Hence there is at least one "good" value of q.

Pick such a value of q and set $A' = S_{j(q)}$. Then, by what we have discovered so far, the map $\psi = \psi_1 \circ \psi_2(q) \circ \psi_3 \circ \psi_4$ is a Freiman k-isomorphism from A' to a subset of $\mathbb{Z}_m$. □

Corollary 25 *Let $A \subseteq \mathbb{Z}$ have cardinality n and suppose that $|A| = n$ and that $|A+A| \leq Cn$. Then $2A - 2A$ contains a proper arithmetic progression of dimension at most $2^{11}C\log C$ and size at least $\exp\left(-2^{16}C(\log C)^2\right)n$.*

Proof. This results from combining Proposition 24, Theorem 11 and Proposition 21. Take $k = 8$ in Proposition 24 and choose a prime $m \in (C^{16}n, 2C^{16}n]$. Then we know that there is $A' \subseteq A$, $|A'| \geq n/8$, which is 8-isomorphic to a subset of $\mathbb{Z}_m$. Let X be this subset; then $|X| \geq m/16C^{16}$. Furthermore it is easy to see that $|X+X| = |A'+A'|$ (this is a consequence of the fact that $A' \cong_8 X$) and so

$$|X+X| \;=\; |A'+A'| \;\leq\; |A+A| \;\leq\; C|A| \;\leq\; 8C|A'| \;=\; 8C|X|.$$

Hence, by Theorem 11, $2X - 2X$ contains a Bohr neighbourhood $B(K, \delta)$ where $|K| \leq 2^{11}C\log C$ and $\delta \geq (2^{15}C\log C)^{-1}$. By Proposition 21 this means that $2X - 2X$ contains a proper (multidimensional) AP of dimension at most $2^{11}C\log C$ and size at least $\exp\left(-2^{16}C(\log C)^2\right)n$.

Now the fact that $A' \cong_8 X$ implies that $2A' - 2A' \cong_2 2X - 2X$, and the property of being a proper multidimensional AP of a given size and dimension is preserved under 2-isomorphism (these are all easy checks). Hence $2A' - 2A'$ also contains a large AP, and hence, *a fortiori*, so does $2A - 2A$. □

3.3 The final argument.

The following proposition concludes the proof of Freiman's theorem.

Proposition 26 (Chang) *Suppose that $A \subseteq \mathbb{Z}$ has size n, that $|A + A| \leq Cn$ and that $2A - 2A$ contains a proper progression P of size ηn and dimension d. Then A is contained in a progression of size at most $2^d(C^4\eta^{-1})^{5C}n$ and dimension at most $d + 4C\log(C^4/\eta)$.*

Proof. We describe an algorithm for selecting some non-negative integer t and subsets S_i, $i \leq t$, of A. Set $P_0 = P$. Let R_0 be a maximal subset of A for which the translates $P_0 + x$, $x \in R_0$, are all disjoint. If $|R_0| \leq 2C$ then set $t = 0$ and $S_0 = R_0$, and terminate the algorithm. Otherwise take S_0 to be any subset of R_0 of cardinality $2C$, and set $P_1 = P_0 + S_0$. Take R_1 to be a maximal subset of A for which the translates $P_1 + x$, $x \in R_1$, are all distinct. If $|R_1| \leq 2C$ then set $t = 1$ and $S_1 = R_1$ and terminate the algorithm. Otherwise choose $S_1 \subseteq R_1$ with $|S_1| = 2C$ and set $P_2 = P_1 + S_1$. Continue in this way.

We claim that this is a finite algorithm, and that in fact $t \leq \log(C^4/\eta)$. Indeed the fact that the translates $P_i + x$, $x \in S_i$, are all disjoint means that $|P_{i+1}| = |P_i||S_i|$ for $i \leq t-1$. It follows that

$$|P_t| \;\geq\; |P||S_0|\ldots|S_{t-1}| \;\geq\; \eta(2C)^t n. \tag{18}$$

Observe, however, that

$$P_t \;\subseteq\; P + A + A + \cdots + A,$$

where there are t copies of A. Since $P \subseteq 2A - 2A$ this means that $P_t \subseteq (t+2)A - 2A$, and hence by Plünnecke's inequality we have $|P_t| \leq C^{t+4}n$. Comparison with (18) proves the claim.

Let us examine what happens when the algorithm finishes. Then we have a set $R_t \subseteq A$, $|R_t| \leq 2C$, which is maximal subject to the translates $P_t + x$, $x \in R_t$, being disjoint. In other words if $a \in A$ then there is $x \in R_t$ such that $(P_t + a) \cap (P_t + x) \neq \emptyset$, and so

$$A \subseteq P_t - P_t + R_t \subseteq (P - P) + (S_0 - S_0) + \cdots + (S_{t-1} + S_{t-1}) + R_t. \tag{19}$$

If $S = \{s_1, \dots, s_k\}$ is a subset of an abelian group define $\overline{S}$ to be the *cube* spanned by S, that is the set of everything of the form $\sum_i \epsilon_i s_i$ where $\epsilon_i \in \{-1, 0, 1\}$. $\overline{S}$ is a multidimensional progression of dimension $|S|$ and size at most $3^{|S|}$, and it contains the set $S - S$. It follows from (19) that $A \subseteq Q$, where Q is the multidimensional progression

$$Q = P - P + \overline{S}_0 + \cdots + \overline{S}_{t-1} + \overline{R}_t.$$

The dimension of Q satisfies

$$\begin{aligned}\dim(Q) &\leq \dim(P) + \sum_{i=0}^{t-1} |S_i| + |R_t| \\ &\leq d + 2C(t+1) \\ &\leq d + 4C\log(C^4/\eta).\end{aligned}$$

To estimate the size of Q, note that the properness of P implies that $|P - P| = 2^d|P|$. Hence

$$\begin{aligned}|Q| &\leq |P - P| \cdot \prod_{i=0}^{t-1} 3^{|S_i|} \cdot 3^{|R_t|} \\ &\leq 2^d 3^{2C(t+1)}|P| \\ &\leq 2^d 3^{4C\log(C^4/\eta)} C^4 n \\ &\leq 2^d \left(\frac{C^4}{\eta}\right)^{5C} n,\end{aligned}$$

the penultimate step following from Plünnecke's inequality and the fact that $P \subseteq 2A - 2A$.□

Combining this proposition with Corollary 25 gives the following effective version of Freiman's theorem. The bounds here are, except for the constants, the best known at the present time.

Theorem 27 (Effective Freiman Theorem) *Let $A \subseteq \mathbb{Z}$ have cardinality n and suppose that $|A + A| \leq Cn$. Then A is contained in a multidimensional AP of dimension at most $2^{20}C^2(\log C)^2$ and size at most $\exp(2^{20}C^2(\log C)^2)\, n$.*

3.4 Applications of Freiman's theorem.

The most substantial application of Freiman's theorem that I know of occurs in W.T. Gowers' proof of Szemerédi's theorem with explicit bounds.

Theorem 28 (Gowers) *Let $k \geq 3$ be an integer and let A be a subset of $\{1, \ldots, N\}$ of density at least $(\log \log N)^{-2^{-2^{k+9}}}$. Then A contains a non-trivial arithmetic progression of length k.*

I cannot even begin, in these lectures, to give an outline of why Freiman's theorem should be relevant to this problem. The reader is referred to [5]. Although this paper is some 123 pages long, it is not such a great effort to read the proof of the case $k = 4$. If the reader does this she will encounter several ideas related to those we have discussed here.

There are many applications of the various intermediate results we have proved. For another application of Chang's structure theorem, see [9].

3.5 Further remarks.

One reasonably natural question to ask is whether we can ensure that the progression obtained in Freiman's theorem be proper. It turns out that we can, using the following unpublished result of Gowers and Walters:

Theorem 29 (Gowers – Walters) *Let P be an AP of dimension d. Then there is a proper progression P' with $P \subseteq P'$ and $|P'| \leq d^{d^3}|P|$.*

The application of this result to Freiman's theorem does result in a rather bigger progression (of size more like $e^{C^{6+\epsilon}}n$). Walters has suggested that he can replace the exponent 3 in Theorem 29 with 2, which would result in a bound of the form $e^{C^{4+\epsilon}}n$ in Freiman.

Another very natural question to ask is what the best bounds one could hope for might be. For applications (such as modifying Gowers' arguments on Szemerédi's theorem to cover arithmetic progressions in the primes) it would be of great interest if one could take the dimension in Freiman's theorem to be C^{ϵ}. Unfortunately an easy example shows that this is too much to hope for. Indeed take $A = I + S$ where $I = [1, \ldots, \lfloor n/C \rfloor]$ and S is a suitably "spread out" set. Then $|A + A| \approx Cn$, but any covering of A by a progression of dimension substantially less than C will be very uneconomical.

This example has the property that large subsets of A, that is each translate of I, are very highly structured indeed. For applications (I am thinking once again of progressions in primes) it would be of the utmost interest to prove a result of the following kind.

Conjecture 30 (Gowers,[6]) *Suppose that $A \subseteq \mathbb{Z}$, that $|A| = n$ and that $|A + A| \leq Cn$. Then there is a reasonably large set $A' \subseteq A$ which is economically contained in a progression of dimension $d = O(C^{\epsilon})$.*

One might even suggest that A can be covered by a small number of small- dimensional APs (I am not going to try and make any of these statements precise).

A key point in the proof of Freiman's theorem was the location of arithmetic progressions in $2A-2A$, where $A \subseteq \mathbb{Z}_N$ was reasonably large and had small sumset. Without any hypothesis on $|A+A|$, Theorems 11 and 21 combine to give

Theorem 31 *Suppose that $A \subseteq \mathbb{Z}_N$ has cardinality αN. Then $2A - 2A$ contains an AP of dimension at most $8\alpha^{-1}\log(1/\alpha)$ and cardinality at least $\exp\left(-C\alpha^{-1}\log(1/\alpha)^2\right) N$, where C is an absolute constant.*

There are no examples to rule out the possibility that the "correct" dimension here should be $C\log(1/\alpha)$. Any bound of the form $\alpha^{-\epsilon}$ would be of similar interest and applicability to a proof of Conjecture 30.

3.6 Acknowledgements.

These notes derive from three main sources. Firstly I have made use of the book [11], for which Nathanson deserves much credit as these beautiful topics were previously not widely known. Secondly I benefitted hugely from attending a course by Gowers at Cambridge in 1999, in which he explained the mathematics underlying his proof of Szemerédi's theorem. Finally I am indebted to Imre Ruzsa for sharing with me his ideas on Chang's Theorem. The vast majority of the material in Lecture 2 is based on conversations I had with him in Budapest.

References

[1] Bogolyubov, N. *Sur quelques propriétés arithmétiques des presque-périodes (Ukrainian, French)*, Ann. Chaire Phys. Math. Kiev **4**, (1939). 185–205.

[2] Bollobás, Béla *Modern graph theory*, Graduate Texts in Mathematics, 184. Springer-Verlag, New York, 1998.

[3] Chang, M.C. *A polynomial bound in Freiman's theorem*, to appear in Duke Math. Jour.

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

[5] Gowers, W. T. *A new proof of Szemerédi's theorem*, Geom. Funct. Anal. **11** (2001), no. 3, 465–588.

[6] Gowers, W. T. *Rough structure and classification*, GAFA 2000 (Tel Aviv, 1999). Geom. Funct. Anal. 2000, Special Volume, Part I, 79–117.

[7] Green, B.J. *Some constructions in the inverse spectral theory of cyclic groups*, preprint, available at
`http://www.dpmms.cam.ac.uk/~bjg23/preprints.html`

[8] Green, B.J. *Spectral structure of sets of integers*, written for the proceedings of a conference on Fourier analysis and convexity (Milan 2001). Available at
`http://www.dpmms.cam.ac.uk/~bjg23/preprints.html`

[9] Green, B.J. *Arithmetic progressions in sumsets*, to appear.

[10] Katz, Nets Hawk; Tao, Terence *Bounds on arithmetic projections, and applications to the Kakeya conjecture*, Math. Res. Lett. **6** (1999), no. 5-6, 625–630.

[11] Nathanson, Melvyn B. *Additive number theory. Inverse problems and the geometry of sumsets*, Graduate Texts in Mathematics, 165. Springer-Verlag, New York, 1996.

[12] Plünnecke, H. *Eigenschaften und Abschätzungen von Wirkingsfunktionen*, BMwF-GMD-22 Gesellschaft für Mathematik und Datenverarbeitung, Bonn 1969

[13] Ruzsa, Imre Z. *An analog of Freiman's theorem in groups*, Structure theory of set addition. Astérisque No. **258** (1999), xv, 323- -326.

[14] Ruzsa, I. Z. *Generalized arithmetical progressions and sumsets*, Acta Math. Hungar. **65** (1994), no. 4, 379–388.

[15] Ruzsa, I.Z. *An application of graph theory to additive number theory*, Scientia, Ser. A **3** (1989), 97–109