CHAPTER 3

# Small doubling and the Freiman-Ruzsa theorem

## 3.1. Introduction

In the last section we made a fairly detailed study of the operation of set addition, and we defined the doubling constant $\sigma[A] := |A{+}A|/|A|$ of a set $A$. Colloquially, we say that a set $A$ has *small doubling* if $\sigma[A]$ is rather small: this term is particularly appropriate when $\sigma[A]$ is constant, say $\sigma[A] \sim 100$.

A central question in additive combinatorics, the "inverse question for small doubling", asks for some description of those sets $A$ with small doubling. This question may be asked in any abelian group, or indeed in nonabelian groups as we shall do in a later chapter.

The notion of small doubling is somewhat hereditary: if $\sigma[A] = K$ and if $A' \subseteq A$ is a set of cardinality $\delta|A|$ then it is clear that $\sigma[A'] \leqslant K/\delta$. For this reason it is of particular interest to classify sets of small doubling which are "basic" in that every set of small doubling is economically contained within a basic set.

## 3.2. Small doubling in $\mathbb{F}_2^\infty$

Consider the vector space $\mathbb{F}_2^\infty$ consisting of all infinite sequences $(a_n)_{n=1}^\infty$ with $a_n \in \mathbb{F}_2$. This group is often a very useful *model* setting in which to test additive combinatorial arguments. The huge amount of torsion in this group can be very helpful. I know of no better place to see this than in the following result of Imre Ruzsa, giving an answer to the inverse question for small doubling in $\mathbb{F}_2^\infty$. When thinking about $\mathbb{F}_2^\infty$, recall that addition is the same as subtraction!

Theorem 3.1 (Ruzsa). *Suppose that $A \subseteq \mathbb{F}_2^\infty$ is a finite set with $\sigma[A] \leqslant K$. Then $A$ is contained inside some subspace $H$ with $|H| \leqslant \exp(CK^C)|A|$.*

*Proof.* Pick a set $X \subseteq 3A$ with the property that the translates $A{+}x$, $x \in X$, are all disjoint, and which is maximal with respect to this property. Since $\bigcup_{x\in X}(x + A) \subseteq 4A$, it follows that $|X| \leqslant |4A|/|A|$, a quantity which is bounded by $CK^C$ from Ruzsa calculus. Now suppose that $y \in 3A$. By the supposed maximality of $X$ we must have $(y{+}A)\cap(x{+}A) \neq \emptyset$ for some $x \in X$, which means that $y \in X + 2A$. It follows that $3A \subseteq X + 2A$. Adding copies of $A$, we then obtain $4A \subseteq X + 3A \subseteq 2X + 2A$, $5A \subseteq 3X{+}2A$, and so on. It follows that the group $\langle A\rangle$ generated by $A$ is contained in $\langle X \rangle + 2A$, a set of size at most $2^{|X|}|2A| = 2^{|X|}K|A|$. The theorem follows immediately. $\square$

### 3.3. Small doubling in $\mathbb{Z}$: GAPs and the Freĭman-Ruzsa theorem

The inverse question in $\mathbb{Z}$ is more complicated. For a start, there do not exist sets with more than one element whose doubling constant is 1. Indeed we have the following almost trivial bound.

Lemma 3.1. *Suppose that $A \subseteq \mathbb{Z}$ is a set of $n$ integers. Then $|A + A| \geqslant 2n - 1$.*

*Proof.* Order the elements of $A$ as $a_1 < a_2 < \cdots < a_n$. Then we have

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

and explicit exhibition of $2n - 1$ distinct elements of $A$. $\square$

One can show without too much pain (see the example sheet) that equality occurs if an only if $A$ is an arithmetic progression of length $n$. Since small doubling is hereditary, any reasonably dense subset of an arithmetic progression will have small doubling. A crucial observation, however, is that these are not the only examples.

Definition 3.1 (Generalised Arithmetic Progression or GAP). Suppose that $x_0$ and $x_1, \dots, x_d$ are integers and that $L_1, \dots, L_d$ are positive integers. Then any set of the form

$$P := \{x_0 + l_1 x_1 + \cdots + l_d x_d : 0 \leqslant l_i < L_i\}$$

is called a GAP. The dimension of $P$ is declared to be $d$ and its size $L_1 \dots L_d$. $P$ is said to be *proper* if all $L_1 \dots L_d$ elements are distinct, that is to say if $|P| = L_1 \dots L_d$.

Now $P + P$ is contained within the GAP $\{2x_0 + l_1 x_1 + \cdots + l_d x_d : 0 \leqslant l_i < 2L_i - 1\}$, and so in particular if $P$ is proper then $\sigma[P] \leqslant 2^d$. Thus these GAPs must be added to our list of "basic" examples.

Remarkably there are no further examples. This is the content of the so-called *Freĭman-Ruzsa Theorem.* We will prove a version of this theorem with relatively good bounds; this result is due to Chang [**?**] though much of the proof is based on Ruzsa's ideas [**?**].

Theorem 3.2 (Freĭman-Ruzsa; Chang). *Suppose that $A \subseteq \mathbb{Z}$ is a finite set and that $\sigma[A] \leqslant K$. Then there is a* GAP*, $P$, with* $\dim(P) \leqslant CK^C$ *and* $|P| \leqslant \exp(CK^C)|A|$*, which contains $A$.*

### 3.4. Freĭman homomorphisms

In his remarkably insightful 1966 book [**?**], Freĭman made an attempt to treat additive number theory (as it was then) by analogy with the way Klein treated geometry: as well as sets $A, B, \dots$ 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 Freǐman homomorphism is now quite important in additive combinatorics.

Definition 3.2 (Freǐman homomorphism). Suppose that $s \geqslant 2$ is an integer. Suppose that $A$ is a subset of some abelian group $G$ and that $B$ is a subset of some other abelian group $H$. Then we say that a map $\phi : A \rightarrow B$ is a Freiman $s$-homomorphism if we have

$$\phi(a_1) + \cdots + \phi(a_s) = \phi(a'_1) + \cdots + \phi(a'_s)$$

whenever $a_1 + \cdots + a_s = a'_1 + \cdots + a'_s$.

If $\pi : G \rightarrow H$ is a group homomorphism then $\pi$ induces a Freiman homomorphism of all orders on any set $A \subseteq G$. The notion of Freiman homomorphism is rather more general, however. For example, an arbitrary bijection between $\{1, 10, 100, 1000\}$ and $\{1, 100, 10000, 1000000\}$ is a Freiman 2-homomorphism since neither of these sets has any nontrivial additive relations $a + b = c + d$. We shall see some more useful examples shortly.

The map $\phi$ is said to be a Freiman $s$-isomorphism if it has an inverse $\phi^{-1}$ which is also a Freǐman $s$-homomorphism. We caution that, contrary to what is often expected in more algenraic situations, a 1-1 Freiman homomorphism need not be a Freiman isomorphism. An excellent example of this is the obvious map

$$\phi : \{0, 1\}^n \rightarrow \mathbb{F}_2^n$$

which is a Freiman homomorphism of all orders. It is not, however, a Freiman isomorphism of any order since there are a great many more additive relations amonst elements of $\mathbb{F}_2^n$ than there are in $\{0, 1\}^n$.

In the following lemma we record some more-or-less easy to prove facts about Freiman homomorphisms.

Lemma 3.2 (Basic facts about Freiman homomorphisms). *Suppose that $A, B$ and $C$ are sets inside abelian groups. Let $s \geqslant 2$ be an integer. We have the following.*

(i) *Suppose that $\phi : A \rightarrow B$ and $\psi : B \rightarrow C$ are Freiman $s$-homomorpisms. Then so is the composition $\psi \circ \phi$.*
(ii) *Suppose that $\phi : A \rightarrow B$ 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 $\phi : A \rightarrow B$ is a Freiman $s$-homomorphism and let $k, l \geqslant 0$ be integers. Then $\phi$ induces a Freiman $s'$ homomorphism $\tilde{\phi} : kA - lA \rightarrow kB - lB$, for any integer $s' \leqslant s/(|k| + |l|)$.*
(iv) *The above three statements are true with "homo" replaced by "iso" throughout.*
(v) *Suppose that $P$ is a* GAP *and that $\phi : P \rightarrow B$ is a Freiman $s$-homomorphism for some $s \geqslant 2$. Then $\phi(P)$ is a* GAP *of the same dimension.*
(vi) *Suppose that $m \geqslant 1$ is an integer, and consider the group $\mathbb{Z}/m\mathbb{Z}$ together with the "unwrapping" map $\psi : \mathbb{Z}/m\mathbb{Z} \rightarrow \{1, \dots, m\}$ which sends a residue $x$ to the least positive integer congruent to $x (\operatorname{mod} m)$. Suppose*

*that $A \subseteq \mathbb{Z}/m\mathbb{Z}$ and that $\psi(A)$ is contained in an interval of length at most $m/s$. Then $\psi$ is a Freiman $s$-homomorphism on $A$.*

*Proof.* Essentially all of this is quite easy, and is a good exercise in checking that one has understood the definitions. (i) is trivial. (ii) may be established by the simple expedient of introducing dummy variables. Indeed if

$$a_1 + \cdots + a_{s'} = a'_1 + \cdots + a'_{s'}$$

then we may add $s - s'$ copies of $a$ to each side, for some arbitrary $a \in A$. Since $\phi$ is a Freiman $s$-homomorphism we then have

$$\phi(a_1) + \cdots + \phi(a_{s'}) + \phi(a) + \cdots + \phi(a) = \phi(a'_1) + \cdots + \phi(a'_{s'}) + \phi(a) + \cdots + \phi(a);$$

cancelling the $\phi(a)$s from both sides yields the result. To prove (ii), define $\tilde{\phi} : kA - lA \rightarrow H$ by

$$\tilde{\phi}(a_1 + \cdots + a_k - a'_1 - \cdots - a'_l) = \phi(a_1) + \cdots + \phi(a_k) - \phi(a'_1) - \cdots - \phi(a'_l).$$

It is conceptually easy, though notationally rather irritating, to verify that $\tilde{\phi}$ is well-defined and in fact defines a Freiman homomorphism of the order stated.

Part (iv) is immediate: simply apply (ii) and (iii) to the inverse map $\phi^{-1}$.

To prove (v) it clearly suffices to assume that $s = 2$. Let $\phi : P \rightarrow \phi(P)$ be a Freiman 2-isomorphism, and 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 $y_0 + y_1 = \phi(x_0 + x_i)$ for $i = 0, 1, \dots, d$; 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 $l_1 + \cdots + l_d$, noting that we have defined the $y_i$ in such a way that it holds whenever $l_1 + \cdots + l_d = 0$ 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 let us consider (vi). The point here is that if one restricts attention to sets $A \subseteq \mathbb{Z}/m\mathbb{Z}$ for which $\psi(A)$ is contained in an interval of length at most $m/s$ then two sums $a_1 + \cdots + a_s$ and $a'_1 + \cdots + a'_s$ are equal mod $m$ if and only if their lifts $\psi(a_1) + \cdots + \psi(a_s)$ and $\psi(a'_1) + \cdots + \psi(a'_s)$ are equal, since there is no "wraparound". □

### 3.5. Ruzsa's model lemma

In this section we prove a remarkable lemma of Imre Ruzsa. It asserts that a subset of $\mathbb{Z}$ with small doubling has a large piece which is Freiman isomorphic to a dense subset of a cyclic group $\mathbb{Z}/m\mathbb{Z}$. In that setting one has tools available which are cannot be brought to bear on subsets of $\mathbb{Z}$ which, despite having small doubling, might be highly "spread out". Most particularly one may use harmonic analysis as an effective tool in $\mathbb{Z}/m\mathbb{Z}$.

Proposition 3.1 (Ruzsa's model lemma). *Suppose that $A \subseteq \mathbb{Z}$ 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 $A' \subseteq A$ with $|A'| \geqslant |A|/s$ which is Freiman $s$-isomorphic to a subset of $\mathbb{Z}/m\mathbb{Z}$.*

*Proof.* By translating $A$ if necessary we may assume that $A$ consists of positive integers. Let $q$ be a prime number greater than all elements of $A$, and consider the composition of maps

$$\mathbb{Z} \xrightarrow{\pi_q} \mathbb{Z}/q\mathbb{Z} \xrightarrow{D_\lambda} \mathbb{Z}/q\mathbb{Z} \xrightarrow{\psi} \mathbb{Z} \xrightarrow{\pi_m} \mathbb{Z}/m\mathbb{Z}$$

where $\pi_q$ and $\pi_m$ are reduction mod $q$ and $m$ respectively, $D_\lambda$ is multiplication (dilation) by $\lambda \in (\mathbb{Z}/q\mathbb{Z})^\times$ and $\psi$ is the unfolding map from $\mathbb{Z}/q\mathbb{Z}$ to $\{1, \dots, q\}$.

Now $\psi_q, D_\lambda$ and $\pi_m$ are Freiman homomorphisms of any order. By Proposition **??** (vi) $\psi$ is a homomorphism of order $s$ when restricted to any subset of $\mathbb{Z}/q\mathbb{Z}$ whose unfolding lies in a subinterval of $\mathbb{Z}$ of length at most $q/s$. Since $\mathbb{Z}/q\mathbb{Z}$ may be partitioned into $s$ sets with this property (the inverse images under $\psi$ of the intervals $\{x \in \mathbb{Z} : jq/s < x \leqslant (j+1)q/s\}$), it follows from the pigeonhole principle that for each $\lambda$ there is a set $A' \subseteq A$, $|A'| \geqslant |A|/s$, with the property that the composition $\phi := \pi_m \circ \psi \circ D_\lambda \circ \pi_q$ is a Freiman $s$-homomorphism when restricted to $A'$.

Everything we have said so far holds for an arbitrary $\lambda$. To conclude the proof we show that there is a choice of $\lambda$ for which $\phi|_A$ is invertible, and for which its inverse is also a $s$-homomorphism. For this choice of $\lambda$, $\phi$ will then be a Freiman $s$-isomorphism when restricted to the set $A'$ just defined. To this end it suffices to show that whenever

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

we have

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

since this clearly implies that $\phi$ is one-to-one. The only way in which these conditions can fail to hold, for a given $\lambda$, is if there is some non-zero expression $d = a_1 + \cdots + a_s - a_1' - \cdots - a_s'$ such that $\pi_m \circ \psi \circ \pi_q(\lambda d) = 0$. Let us fix $d$ and ask about values of $\lambda$ for which this phenomenon occurs: lacking imagination, we call them "bad for $d$". As $\lambda$ ranges over $(\mathbb{Z}/q\mathbb{Z})^\times$, $\pi_q(\lambda d)$ of course covers $(\mathbb{Z}/q\mathbb{Z})^\times$ uniformly, and hence the "unwrapped" set $\psi \circ \pi_q(\lambda d)$ covers each point of $\{1, \dots q-1\}$ precisely once. The number of elements $x$ in this interval for which $\pi_m(x) = 0$ (that is to say $x$ is divisible by $m$) is at most $(q-1)/m$. Since each $d$ lies in the set $(sA - sA) \setminus \{0\}$, it follows that the number of $\lambda$ 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 $m \geqslant |sA - sA|$. It follows that there is at least one $\lambda$ which is not bad for any $d$. By our discussion, the map $\phi$ will then have an inverse which is a Freiman homomorphism of order $s$. □

In our applications of this lemma the set $A$ will have small doubling, we shall take $s = 8$, and it will also be convenient to take $m$ to be prime.

Corollary 3.1 (Ruzsa's model lemma). *Suppose that $A \subseteq \mathbb{Z}$ is a finite set with $\sigma[A] \leqslant K$. Then there is a prime $p \leqslant CK^C|A|$ and a subset $A' \subseteq A$ with $|A'| \geqslant |A|/8$ such that $A'$ is Freiman 8-isomorphic to a subset of $\mathbb{Z}/p\mathbb{Z}$.*

*Proof.* It follows from Corollary 2.1 that $|8A - 8A| \leqslant CK^C|A|$ (in fact, if one has the Plünnecke-Ruzsa inequalities mentioned just after that corollary we obtain the more precise bound $|8A - 8A| \leqslant K^{16}|A|$). Now by Bertrand's postulate (or some even weaker result) there is a prime $p$ satisfying $|8A - 8A| \leqslant p \leqslant 2|8A - 8A|$. This prime of course satisfies the bound $p \leqslant CK^C|A|$, and by the preceding proposition there is a subset $A'$ of $A$ with $|A'| \geqslant |A|/8$ which is Freiman 8-isomorphic to a subset of $\mathbb{Z}/p\mathbb{Z}$. □

## 3.6. Bogolyubov's lemma

Suppose that $A \subseteq \mathbb{Z}$ has doubling constant at most $K$. By applying Ruzsa's model lemma, we can locate a set $S \subseteq \mathbb{Z}/p\mathbb{Z}$ with size $\sigma p$, where $\sigma \geqslant cK^{-C}$, which is Freiman 8-isomorphic to a subset of $A$ and, furthermore, for which $|S| \geqslant \frac{1}{8}|A|$. The aim is now to find some structural properties of $S$ which may be "pulled back" under this Freiman 8-isomorphism in order to tell us something about $A$. This is a little trickier than it might at first sight seem: a Freiman 8-isomorphism can only "see" sums of at most 8 terms such as $s_1 \pm \cdots \pm s_8$, and in particular it cannot see all of the group $\mathbb{Z}/p\mathbb{Z}$.

The following result, known as Bogolyubov's lemma, is what we need. Before stating it we require a definition.

Definition 3.3 (Bohr sets in $\mathbb{Z}/p\mathbb{Z}$). Suppose that $R = \{r_1, \dots, r_k\}$ is a set of nonzero elements of $\mathbb{Z}/p\mathbb{Z}$ and that $\varepsilon > 0$ is a parameter. Then we define the Bohr set $B(R, \varepsilon)$ with frequency set $R$ and width $\varepsilon$ by

$$B(R, \varepsilon) := \{x \in \mathbb{Z}/p\mathbb{Z} : \|r_i x/p\|_{\mathbb{R}/\mathbb{Z}} \leqslant \varepsilon \quad \text{for } i = 1, \dots, k\}.$$

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

Lemma 3.3 (Bogolyubov). *Suppose that $S$ is a subset of $\mathbb{Z}/p\mathbb{Z}$ with cardinality $\sigma p$. Then $2S - 2S$ contains a Bohr set of dimension at most $4/\sigma^2$ and width at least $\frac{1}{10}$.*

*Proof.* We use harmonic analysis on $\mathbb{Z}/p\mathbb{Z}$. Thus if $f : \mathbb{Z}/p\mathbb{Z} \to \mathbb{C}$ is a function then, as in Chapter 1, we define $\widehat{f}(r) := \mathbb{E}_{x \in \mathbb{Z}/p\mathbb{Z}} f(x) e(-rx/p)$. We recall Parseval's identity, namely that

$$(\mathbb{E}_x |f(x)|^2)^{1/2} = \|f\|_2 = \|\widehat{f}\|_2 = (\sum_r \widehat{f}(r))^{1/2}.$$

Let us also recall the notion of convolution: if $g : \mathbb{Z}/p\mathbb{Z}$ is another function then we write

$$f * g(x) = \mathbb{E}_y f(y)g(x-y),$$

whereupon we have the relation $(f*g)^\wedge = \widehat{f}\widehat{g}$. In Chapter 1 we saw that $\|f\|_{U^2} = \|f * \|_2^{1/2}$, which was at the time our justification for introducing the convolution. The most basic reason for an additive combinatorialist to be interested in convolving functions is the observation that if $f = 1_U$ and $g = 1_V$ then $f * g$ has support $U + V$. In other words, $1_U * 1_V$ is a function supported on $U + V$ whose Fourier transform is easy to understand in terms of $\widehat{1}_U$ and $\widehat{1}_V$.

We shall also require a further, very important, property of the Fourier transform: the inversion formula. This states that $f$ may be recovered from its Fourier transform via the relation

$$f(x) = \sum_r \widehat{f}(r)e(rx/p).$$

It is very easily proved using the orthogonality relations as in the proof of Lemma 1.4. Note that we are *summing* over $r$ and not taking expectations: see the discussion immediately following the proof of Lemma 1.4 for the reasons for this.

Returning to the proof of Bogolyubov's lemma, note that $f(x) = 1_S * 1_S * 1_{-S} * 1_{-S}(x)$ is supported on $2S - 2S$. Note also that $\widehat{1}_S(r) = \overline{\widehat{1}_{-S}(r)}$, and so $\widehat{f}(r) = |\widehat{1}_S(r)|^4$.

Let $R$ be the set of all $r \neq 0$ for which $|\widehat{1}_S(r)| \geqslant \sigma^{3/2}/2$. Since $\|1_S\|_2 = \sigma^{1/2}$, it follows immediately from Parseval's identity that $|R| \leqslant 4/\sigma^2$. We claim that $B(R, \frac{1}{10}) \subseteq 2S - 2S$. To prove this, it suffices by our earlier discussion to show that $f(x) > 0$ whenever $x \in B(R, \varepsilon)$.

By the inversion formula and the fact that $f$ is real we have

$$f(x) = \sum_r |\widehat{1}_S(r)|^4 e(rx/p) = \sum_r |\widehat{1}_S(r)|^4 \cos(2\pi rx/p).$$

To bound this sum below, we divide it into three pieces: the term $r = 0$, the terms with$r \in R$, and all other terms. Clearly

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

Now if $r \in R$ then $\cos(2\pi rx/p) \geqslant 0$, and so we simply bound the sum of these terms below by 0. Finally

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

and this is equal to $\sigma^4/4$ by Parseval's identity. Combining all of this we obtain

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

as required. □

## 3.7. Geometry of numbers and progressions in Bohr sets

The ultimate aim of this section is to investigate the structure of Bohr sets, and in particular to establish that they contain large GAPs. In order to do this we must first develop some tools from a a subject called the geometry of numbers.

In so far as we shall investigate it in this course, this is the study of lattice points inside convex bodies. For us, $K \subseteq \mathbb{R}^d$ will be a centrally symmetric convex body, that is to say a set such that $-\mathbf{x}$ and $\frac{1}{2}(\mathbf{x}+\mathbf{y})$ lie in $K$ whenever $\mathbf{x}$ and $\mathbf{y}$ do. The closed centrally symmetric convex bodies are exaactly the same thing as the unit balls of norms on $\mathbb{R}^d$, and let us recall that this is often a useful way of thinking about them.

A *lattice* is a discrete subgroup of $\mathbb{R}^d$. We shall assume some of the basic theory of lattices as may be found in any number of books. In particular any lattice $\Lambda$ whose $\mathbb{R}$-span is all of $\mathbb{R}^d$ has an *integral basis* $\mathbf{v}_1, \dots, \mathbf{v}_d$, which means that $\Lambda$ is the direct sum $\mathbb{Z}\mathbf{v}_1 \oplus \mathbb{Z}\mathbf{v}_2 \oplus \cdots \oplus \mathbb{Z}\mathbf{v}_d$, or in other words every element of $\Lambda$ can be written using integer coordinates relative to the basis $\mathbf{v}_1, \dots, \mathbf{v}_d$. The *determinant* $\det(\Lambda)$ is defined to be the determinant of the column matrix $(\mathbf{v}_1, \dots, \mathbf{v}_d)$, which turns out not to depend on the particular choice of integral basis, or equivalently the volume of the *fundamental parallelepiped* spanned by $\mathbf{v}_1, \dots, \mathbf{v}_d$.

The following lemma is simple but surprisingly powerful. We say that $\Lambda$ is nondegenerate if $\det(\Lambda) \neq 0$, or equivalently if the $\mathbb{R}$-span of $\Lambda$ is all of $\mathbb{R}^d$.

Lemma 3.4 (Blichfeldt's Lemma). *Suppose that $\Lambda$ is a nondegenerate lattice and that $K$ is a set with $\operatorname{vol}(K) > \det(\Lambda)$. Then there are two distinct points $\mathbf{x}, \mathbf{y} \in K$ with $\mathbf{x} - \mathbf{y} \in \Lambda$.*

*Proof.* This is what is often called a "volume-packing argument". 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 $\Lambda$, or in other words

$$\sum_{\mathbf{x}} 1_K(\mathbf{x} - \mathbf{t}) 1_\Lambda(\mathbf{x}) \leqslant 1$$

for all $\mathbf{t} \in \mathbb{R}^d$. Let $R'$ be much bigger than $R$, and average this last inequality over $\mathbf{t}$ lying in the ball $B(0, R')$ to obtain

$$\sum_{\mathbf{x}} 1_\Lambda(\mathbf{x}) \Big( \frac{1}{\operatorname{vol}(B(0, R'))} \int_{B(0,R')} 1_K(\mathbf{x} - \mathbf{t}) \, d\mathbf{t} \Big) \leqslant 1.$$

Since $K \subseteq B(0, R)$, the inner integral equals 1 if $\|\mathbf{x}\| \leqslant R' - R$ (and zero if $\|\mathbf{x}\| \geqslant R' + R$). Therefore

$$\sum_{\mathbf{x}} 1_\Lambda(\mathbf{x}) 1_{B(0,R'-R)}(\mathbf{x}) \, d\mathbf{x} \leqslant \operatorname{vol}(B(0, R')). \tag{3.1}$$

However it is "clear" by tiling with fundamental parallelepipeds that

$$\lim_{r \to \infty} \frac{1}{\operatorname{vol}(B(0, r))} \sum_{\mathbf{x}} 1_\Lambda(\mathbf{x}) 1_{B(0,r)}(\mathbf{x}) = \frac{1}{\det(\Lambda)}.$$

Letting $R' \to \infty$ in (3.1) and noting that $\mathrm{vol}(B(0,R'))/\mathrm{vol}(B(0,R'-R)) \to 1$, we obtain a contradiction. □

Suppose now that $K \subseteq \mathbb{R}^d$ is a centrally-symmetric convex body and that $\Lambda$ is a lattice. We define the *successive minima* $\lambda_1, \dots, \lambda_d$ of $K$ with respect to $\Lambda$ by defining $\lambda_k$ to be the infimum of all those $\lambda$ for which the dilate $\lambda K$ contains $k$ linearly independent elements of $\Lambda$. Note in particular that the closure $\lambda_k \overline{K}$ does then contain $k$ independent elements of $\Lambda$. We may use this observation to pick a *directional basis* for $\Lambda$ with respect to $K$; choose $\mathbf{b}_1, \mathbf{b}_2, \dots$ in sequence so that $\mathbf{b}_k \in \lambda_k \overline{K} \cap \Lambda$, and such that the vectors $\mathbf{b}_1, \dots \mathbf{b}_k$ span a $k$-dimensional subspace of $\mathbb{R}^d$.

The directional basis $\mathbf{b}_1, \dots, \mathbf{b}_d$ is manifestly a basis for $\mathbb{R}^d$ consisting of elements of $\Lambda$, but we caution that it need not be an integral basis for $\Lambda$. One of the questions on the second example sheet asks you to find an example.

We turn now to one of the most important results in the subject, with one of the more mysterious proofs I have come across.

Theorem 3.3 (Minkowski's second theorem). *Suppose that $K \subseteq \mathbb{R}^d$ is a centrally symmetric convex body and that $\Lambda$ is a nondegenerate lattice. Let the successive minima of $K$ with respect to $\Lambda$ be $\lambda_1, \dots, \lambda_d$. Then we have the inequality $\lambda_1 \dots \lambda_d \,\mathrm{vol}(K) \leqslant 2^d \det(\Lambda)$.*

*Proof.* It is convenient to assume, by passing from $K$ to its interior $K^\circ$ if necessary, that $K$ is *open*. This does not affect any aspect of the statement of the theorem. Fix a directional basis $\mathbf{b}_1, \dots, \mathbf{b}_d$ for $\Lambda$ with respect to $K$. The openness of $K$ has the nice consequence that $\lambda_k K \cap \Lambda$ is spanned (over $\mathbb{R}$) by the vectors $\mathbf{b}_1, \dots, \mathbf{b}_{k-1}$. Indeed if it were not we could choose some further vector $\mathbf{b}$ in $\lambda_k K \cap \Lambda$, and by the openness this would in fact lie in $(\lambda_k - \varepsilon)K \cap \Lambda$ for some $\varepsilon > 0$, contrary to the definition of the successive minima $\lambda_k$.

Write each given $\mathbf{x}$ in coordinates relative to the basis vectors $\mathbf{b}_i$ as $x_1\mathbf{b}_1 + \dots + x_d\mathbf{b}_d$. 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 $\mathbf{b}_1, \dots, \mathbf{b}_{j-1}$ (for $j = 1$, $\phi_1(\mathbf{x}) = \mathbf{x}$). Next, we define a map $\phi : K \to \mathbb{R}^d$ 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 $\phi_j$ and $\phi$. In coordinates we have $\phi_j(\mathbf{x}) = \sum_i c_{ij}(\mathbf{x}) b_i$, where $c_{ij}(\mathbf{x}) = x_i$ for $i \geqslant j$, and $c_{ij}$ depends only on $x_j, \dots, x_d$ for $i < j$. It follows that

$$\phi(\mathbf{x}) = \sum_{i=1}^{d} b_i \left(\lambda_i x_i + \psi_j\left(x_{i+1}, \dots, x_d\right)\right)$$

for certain continuous functions $\psi_j$. It follows easily that $\mathrm{vol}(\phi(K)) = \lambda_1 \dots \lambda_d \,\mathrm{vol}(K)$, the determinant of the Jacobian of the transformation $x'_i = \lambda_i x_i + \psi_i(x_{i+1}, \dots, x_d)$ being particularly easy to evaluate due to the matrix being upper triangular.

Suppose, as a hypothesis for contradiction, that $\lambda_1 \dots \lambda_d \,\mathrm{vol}(K) > 2^d \det(\Lambda)$. By Blichfeldt's lemma and the preceding observation 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 $\mathbf{x} = \sum x_i b_i$ and $\mathbf{y} = \sum y_i b_i$, and suppose that $k$ is the largest index such that $x_k \neq y_k$. Then we have $\phi_i(\mathbf{x}) = \phi_i(\mathbf{y})$ for $i > k$, so that

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

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 $b_k$ 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 non-zero, which is contrary to our earlier observation that $\Lambda \cap \lambda_k K$ is spanned by $\mathbf{b}_1, \dots, \mathbf{b}_{k-1}$. □

Our foray into the geometry of numbers had a specific purpose, namely to clarify the structure of Bohr sets. To conclude this section we record the following corollary in that vein. This will be the only result we need in subsequent work.

Corollary 3.2 (Structure of Bohr sets). *Suppose that $R = \{r_1, \dots, r_k\} \subseteq \mathbb{Z}/p\mathbb{Z}$ is a set of $k$ frequencies, and that $\varepsilon < 1/2$ is a parameter. Then the Bohr set $B(R, \varepsilon)$ contains a proper* GAP *of dimension $k$ and size at least $(\varepsilon/k)^k p$.*

*Proof.* Consider the lattice $\Lambda = p\mathbb{Z}^k + (r_1, \dots, r_k)\mathbb{Z}$. Since $p$ is prime, this may be written as a *direct* sum $p\mathbb{Z}^k \oplus \{0, 1, \dots, p-1\} \cdot (r_1, \dots, r_k)$, a presentation which makes it fairly easy to see that $\det(\Lambda) = p^{k-1}$. Let $K \subseteq \mathbb{R}^k$ be the box $\{\mathbf{x} : \|\mathbf{x}\|_\infty \leqslant \varepsilon\}$, that is to say the set of all $\mathbf{x}$ with $|x_1|, \dots, |x_k| \leqslant \varepsilon$. This is a closed, centrally symmetric, convex body. Let $\mathbf{b}_1, \dots, \mathbf{b}_k$ be a directional basis for $\Lambda$, and let $\lambda_1, \dots, \lambda_k$ be the successive minima. From the basic definitions of these objects we know that $\|\mathbf{b}_i\|_\infty \leqslant \varepsilon\lambda_i$ for all $i = 1, \dots, k$. Set $L_i := \lceil 1/\lambda_i k \rceil$ for $i = 1, \dots, k$. Then if $0 \leqslant l_i < L_i$ we have $\|l_i \mathbf{b}_i\|_\infty \leqslant \varepsilon/k$ and whence

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

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

$$\| \frac{(l_1 x_1 + \dots + l_k x_k) r_i}{p} \|_{\mathbb{R}/\mathbb{Z}} \leqslant \varepsilon$$

for each $i$, or in other words the GAP $\{l_1x_1 + \cdots + l_kx_k : 0 \leqslant l_i < L_i\}$ 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 $k^{-k}(\lambda_1 \dots \lambda_k)^{-1}$, which, by Minkowski's Second Theorem and the fact that $\det(\Lambda) = p^{k-1}$ and $\mathrm{vol}(K) = (2\varepsilon)^k$, is at least $(\varepsilon/k)^k p$.

To establish the properness, suppose that

$$l_1x_1 + \cdots + l_kx_k = l_1'x_1 + \cdots + l_k'x_k (\mathrm{mod}\, p),$$

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 $p\mathbb{Z}^k$ and furthermore

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

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. □

## 3.8. Chang's covering argument and the conclusion of the proof

Ruzsa's model lemma, Bogolyubov's lemma and Corollary 3.2 allow us to proceed from a set $A \subseteq \mathbb{Z}$ with $\sigma[A] \leqslant K$ to the conclusion that $2A - 2A$ contains a large GAP. We will go over the details a little later, but let us first supply the final piece in the proof of the Freĭman-Ruzsa theorem, a covering lemma of Chang which allows us to use this information to efficiently place $A$ *inside* a GAP.

Lemma 3.5 (Chang). *Let $K \geqslant 2$ and that $\eta \leqslant 1/2$. Suppose that $A$ is a finite subset of $\mathbb{Z}$ with $\sigma[A] \leqslant K$ and that $2A - 2A$ contains a proper* GAP *$P$ of size $\eta|A|$ and dimension $d$. Then $A$ is contained in a* GAP *of size at most $2^d \eta^{-CK^C}|A|$ and dimension at most $d + CK^C \log(1/\eta)$.*

*Proof.* We describe an algorithm for selecting some non-negative integer $t$ and subsets $S_i$, $i \leqslant t$, of $A$. Let $L$ be a positive integer to be specified later. 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 $|R_0| \leqslant L$ 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 $L$, 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 $|R_1| \leqslant L$ then set $t = 1$ and $S_1 = R_1$ and terminate the algorithm. Otherwise choose $S_1 \subseteq R_1$ with $|S_1| = L$ and set $P_2 = P_1 + S_1$. Continue in this way.

We claim that for a suitably chosen $L$ this is a finite algorithm. 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

$i \leqslant t-1$. It follows that

$$|P_t| \geqslant |P||S_0|\dots|S_{t-1}| \geqslant \eta L^t n. \tag{3.2}$$

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 Corollary 2.1, the bounded for iterated sumsets under the assumption of small doubling, we have $|P_t| \leqslant K^{C(t+4)}|A|$. If $L = K^{2C}$, then, the algorithm can last no more than $t$ steps where $t \leqslant C\log(1/\eta)$ (this is a slightly crude estimate).

Let us examine what happens when the algorithm finishes. Then we have a set $R_t \subseteq A$, $|R_t| \leqslant L$, 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

$$A \ \subseteq\ P_t - P_t + R_t \ \subseteq\ (P-P) + (S_0 - S_0) + \cdots + (S_{t-1} - S_{t-1}) + R_t. \tag{3.3}$$

Now it is clear that for any finite set $S$ the difference set $S-S$ is contained in a GAP $\overline{S}$ of dimension at most $|S|$ and size at most $3^{|S|}$, namely the one in which the $x_i$ are the elements of $S$ and the "lengths" $L_i$ are all 1. It follows from (3.3) 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

$$\dim(Q) \leqslant \dim(P) + \sum_{i=0}^{t-1} |S_i| + |R_t| \leqslant d + L(t+1) \leqslant d + CK^C\log(1/\eta).$$

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

$$|Q| \leqslant |P-P| \cdot \prod_{i=0}^{t-1} 3^{|S_i|} \cdot 3^{|R_t|} \leqslant 2^d 3^{L(t+1)}|P|.$$

The claimed bound follows immediately from a crude application of the estimate $t \leqslant C\log(1/\eta)$, the choice of $L = K^{2C}$, and the fact that $P \subseteq 2A-2A$ (which means that $|P| \leqslant K^C|A|$ by Corollary 2.1). □

We are now in a position to conclude the proof of the Freĭman-Ruzsa theorem, which is merely a pleasant putting-together of the facts we have just assembled. Suppose that $A \subseteq \mathbb{Z}$ is a set with $\sigma[A] \leqslant K$. Then:

(i) There is a set $A' \subseteq A$ with $|A'| \geqslant |A|/8$ which is Freiman 8-isomorphic to a set $S \subseteq \mathbb{Z}/p\mathbb{Z}$, where $p$ is a prime with $p \leqslant CK^C|A|$ (Ruzsa's model lemma, Corollary 3.1);

(ii) $2S - 2S$ contains a Bohr set $B(R, \frac{1}{10})$ with $|R| \leqslant CK^C$ (Bogolyubov's lemma, Lemma 3.3);

(iii) That Bohr set in turn contains a proper GAP, $P$, with dimension at most $CK^C$ and size at least $\exp(-CK^C)|A|$ (Geometry of numbers, Corollary 3.2);

(iv) The set $2A - 2A$ contains a proper GAP, $\tilde{P}$, with dimension at most $CK^C$ and size at least $\exp(-CK^C)|A|$ (Basic facts about Freiman homomorphisms, specifically Proposition 3.2 (iii), (iv) and (v)), and

(v) $A$ is contained in a GAP with dimension at most $CK^C$ and size at most $\exp(CK^C)|A|$ (Chang's covering lemma, Lemma 3.5).

This concludes the proof of the Freĭman-Ruzsa theorem: it is one of the classics of this or any other subject. □