Sumsets and structure

Imre Z. Ruzsa

Alfréd Rényi Institute of Mathematics, Budapest, Pf. 127, H-1364 Hungary,
E-mail address: ruzsa@renyi.hu

Contents

Foreword

This is a somewhat extended version of my course given in the Doccourse in Barcelona, Spring 2008.

Two students, Itziar Bardaji Goikoetxea and Lluís Vena helped me to prepare the final version, and here I wish to express my sincere thanks to them.

Overview

This booklet is devoted to certain aspects of combinatorial number theory, or additive combinatorics as it is now often called. This change of terminology reflects a shift in the emphasis of problems investigated. First it was mainly infinite sequences and finite sets of integers; this naturally led to sets of residues, then sets in finite groups, and also sets of lattice points, then sets in general commutative groups. (We shall now and then mention results that do not need commutativity, but will not pursue this aim forcefully.)

In classical additive number theory we start with a given set, say of primes, and try to understand how an integer can be expressed as a sum of elements of this set. In combinatorial (or structural, or inverse) theory we do the opposite: given an additive assumption about a set, say that it has few or many sums, we try to understand its structure. (This is not always explicit in the formulation; we can equivalently say "if a set has few sums, it has property A" or "if a set has property non-A, it has many sums"; we will not attempt uniformity here.)

Historically, combinatorial additive theory grew out of the classical. Though a few isolated results existed before, the turning point is Schnirelmann's approach to the Goldbach problem. Goldbach's conjecture asserts that any integer > 3 can be expressed as a sum of 2 or 3 primes, depending on parity. Schnirelmann proved the weaker result that there is a bound k so that every integer is a sum of at most k primes, or in other words, the primes form an additive basis. To this end he established that (very loosely speaking; exact formulations will be given in Chapter 3) integers that can be written as a sum of two primes have positive density; and every set having positive density is a basis. For the Goldbach problem Schnirelmann's approach was soon superseded by Vinogradov's trigonometric sum method; however, it kindled the interest in addition of general sets.

In the first chapter we consider questions of the following kind. Suppose we know the cardinality of a (finite) set and we know also the number of sums of pairs. What can we say about the number of differences, or of sums of triples? The understanding of such cardinality problems is of paramount importance for understanding the structure.

First we explain the most important tool, Plünnecke's inequality, then two further inequalities independent of it. These will be applied to study the connection between $|A|$, $|A + B|$ and $|A + kB|$, with particular emphasis on the case $B = A$.

In the second chapter we prove Freiman's structure theorem.

In the third chapter we tell results connecting geometrical position or position within a group and cardinality of sumsets.

In the fourth chapter we give some results about density.

In the fifth chapter we explore some connections with topology and measure.

Here I express my sincere thanks to

Notation

Let A and B be sets in a (mostly commutative) group. We will call the group operation addition and use additive notation. The *sumset* of these sets is

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

Similarly

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

For repeated addition we write

$$kA = A + \cdots + A,\ k \text{ times};$$

in particular, $1A = A$, $0A = \{0\}$.

The set kA is typically different from

$$k \cdot A = \{ka : a \in A\};$$

this will appear rarely.[1]This should cause no difficulty in a country where ortoraphy distinguishes between -ll– and –l·l–.

We will use $\mathbb{Z}_q = \mathbb{Z}/(q\mathbb{Z})$ to denote the set of residue classes modulo q.

Exercises and problems

The book contains exercises and problems; the difference between them is that an exercise is what I can solve. Some exercises are called a *prexercise*; this means that the solution will be included in the text, just I feel it may benefit the reader to meditate on it at that point before (or better, instead) reading the proof.

The next collection of exercises is deliberately vague. There are obvious answers (for instance, $|A - A| < n^2$ in the next), not obvious but doable which will be told later, and finding the exact bound would be an important new result.

Exercise 1. Let $|A| = n$ and assume that $|2A| \le \alpha n$. Find a bound for $|A - A|$.

Exercise 2. Let $|A| = n$ and assume that $|2A| \le \alpha n$. Find a bound for $|3A|$.

Exercise 3. Let $|A| = n$ and assume that $|A - A| \le \alpha n$. Find a bound for $|2A|$.

1∗

CHAPTER 1

Cardinality inequalities

1. Introduction

Let A, B be sets in a group, $|A| = m$, $|B| = n$. The cardinality of $A + B$ can be anywhere between $\max(m, n)$ and mn. Our aim is to understand the connection between this size and the structure of these sets.

EXERCISE 4. If $A, B \subset \mathbb{Z}$, $|A| = m$, $|B| = n$, prove that

$$|A + B| \geq m + n - 1$$

and describe the cases of equality.

EXERCISE 5. Given three positive integers m, n, s such that $m + n - 1 \leq s \leq mn$, find sets $A, B \subset \mathbb{Z}$ such that $|A| = m$, $|B| = n$, $|A + B| = s$.

In this chapter we present some inequalities of the following kind: if a sumset, say $A + B$ is small (in various senses), then so are some other sums. The most frequently applied one of them sounds as follows.

THEOREM 1.1. *Let A, B be finite sets in a commutative group and write $|A| = m$, $|A + B| = \alpha m$. For arbitrary nonnegative integers k, l we have*

$$|kB - lB| \leq \alpha^{k+l} m.$$

Observe that there is no a priori assumption on the size of B; however, with such assumptions sometimes the conclusion can be strengthened.

We end this introduction by mentioning some basic ideas.

(i) **Direct product.** Assume A_1, A_2, $\dots A_k$ are subsets of a group G with cardinalities of sumsets

$$|A_{i_1} + A_{i_2} + \cdots + A_{i_m}| = N(i_1, \dots, i_m).$$

Let $A_1', \dots$ be another collection of sets in another group G' with corresponding values $N'(\dots)$. If we form the direct products

$$B_i = A_i \times A_i' = \{(a, b) : a \in A, b \in B\} \subset G \times G',$$

then we have

$$|B_{i_1} + B_{i_2} + \cdots + B_{i_m}| = N(i_1, \dots, i_m) N'(i_1, \dots, i_m).$$

This explains the multiplicative nature of many of the results – when a quantity is estimated in terms of others, this is mostly in the form of a product of powers. This method can often be used to build large examples starting from a single one. It will be used now and then in the opposite way: we apply a result for a power of a small set to get better results for the small set (see Section 6).

(ii) **Projection.** If we start from sets of integers, the above construction gives us sets of integral vectors. This is, however, not an essential difference. If we have sets

$A_i \subset \mathbb{Z}^d$ and a *finite* number of sum-cardinalities are prescribed, then we can construct sets of integers that behave the same way. Indeed, the linear map

$$(x_1, \dots, x_d) \to x_1 + mx_2 + \cdots + m^{d-1}x_d$$

will not add any new coincidence between sums if m is large enough.

This observation will be used without any further mentioning. If we construct a set in $\mathbb{Z}^d$ with certain properties, we shall tacitly realize that a set of integers can also be constructed if necessary; a set in several dimensions often exhibits the structure more clearly.

Exercise 6. Extend Exercise 4 to sets in $\mathbb{Z}^d$.

On the other hand if we know that a set is proper d-dimensional, this may yield further results.

Exercise 7. Improve Exercise 4 for sets in $\mathbb{Z}^2$ that do no lie on a single line.

(iii) **Torsion.** The above consideration shows that from our point of view the structure of $\mathbb{Z}^k$ is not richer than that of $\mathbb{Z}$. We can add that no torsionfree group produces anything new either. Indeed, let G be a torsionfree group and take a finite subset (the union of all finite sets which we want to add). This generates a subgroup G'; and, as a finitely generated torsionfree group, G' is isomorphic to $\mathbb{Z}^k$ for some k.

Exercise 8. Extend Exercise 4 to sets in any commutative torsionfree group.

Exercise 9. Extend Exercise 4 to sets in any noncommutative torsionfree group.

2. Plünnecke's method

Plünnecke [**39**] developed a graph-theoretic method to estimate the density of sumsets $A + B$, where A has a positive density and B is a basis. I published a simplified version of his proof [**47, 48**]. Other accounts (of my version) were published by Malouf [**32**] and Nathanson [**33**]. In the sequel we adopt Malouf's terminology.

Plünnecke observed that the cardinality properties of the sets A, $A + B$, $A + 2B$, $\dots$, are well reflected by the following directed graph. We take $h+1$ copies of the group where these sets are situated, and build a graph on these sets as vertices by connecting an $x \in A + jB$ to an $y \in A + (j + 1)B$ if $y = x + b$ with some $b \in B$. We call this graph the *addition graph*. These graphs have certain properties which follow from the commutativity of addition, and hence Plünnecke called them *commutative*; we shall retain this terminology.

We consider directed graphs $\mathcal{G} = (V, E)$, where V is the set of vertices and E is that of the edges. If there is an edge from x to y, then we also write $x \to y$. A graph is *semicommutative*, if for every collection $(x; y; z_1, z_2, \dots, z_k)$ of distinct vertices such that $x \to y$ and $y \to z_i$ there are distinct vertices $y_1, \dots, y_k$ such that $x \to y_i$ and $y_i \to z_i$. $\mathcal{G}$ is *commutative*, if both $\mathcal{G}$ and the graph $\hat{\mathcal{G}}$ obtained by reversing the direction of every edge of $\mathcal{G}$ are semicommutative.

Our graphs will be of a special kind we call *layered*. By an *h–layered* graph we mean a graph with a fixed partition of the set of vertices

$$V = V_0 \cup V_1 \cup \cdots \cup V_h$$

into $h + 1$ disjoint sets (layers) such that every edge goes from some V_{i-1} into V_i.

EXERCISE 10. If there is no isolated point, then this partition is unique.

To avoid the separate formulation of certain degenerate cases we do not exclude isolated points.

For $X, Y \subset V$, we define the *image* of X in Y as

$$\mathrm{im}(X,Y) = \{y \in Y : \text{ there is a directed path from some } x \in X \text{ to } y\}.$$

The *magnification ratio* is defined by

$$\mu(X,Y) = \min\left\{\frac{|\mathrm{im}(Z,Y)|}{|Z|} : Z \subset X, Z \neq \emptyset\right\}.$$

For a layered graph we write

$$\mu_j(\mathcal{G}) = \mu(V_0, V_j).$$

Now Plünnecke's main result can be stated as follows.

THEOREM 2.1 (Plünnecke [**39**]). *In a commutative layered graph $\mu_j^{1/j}$ is decreasing.*

That is, for $j < h$ we have $\mu_h \leq \mu_j^{h/j}$. An obvious (and typically the only available) upper estimate for μ_j is $|V_j|/|V_0|$. This yields the following corollary (in fact, an equivalent assertion).

THEOREM 2.2. *Let $j < h$ be integers, $\mathcal{G}$ a commutative layered graph on the layers $V_0, \dots, V_h$. Write $|V_0| = m$, $|V_j| = s$. There is an $X \subset V_0$, $X \neq \emptyset$ such that*

$$|\mathrm{im}(X, V_h)| \leq (s/m)^{h/j}|X|.$$

EXERCISE 11. Deduce Theorem 2.1 from Theorem 2.2.

These fundamental results will be proved in the next three sections. Now we mention some important corollaries.

An application of the above theorem to the addition graph yields the following result.

THEOREM 2.3. *Let $j < h$ be integers, A, B sets in a commutative group and write $|A| = m$, $|A + jB| = \alpha m$. There is an $X \subset A$, $X \neq \emptyset$ such that*

$$|X + hB| \leq \alpha^{h/j}|X|.$$

It is not true in general that a proper choice for X is A itself. $|A + hB|$ can be much larger, it can be greater than $m^{1+C(h)}$, even if $\alpha < 2$. X has to be selected carefully. For more details on this phenomenon see Section 10.

Since $|X + hB| \geq |hB|$ and $|X| \leq m$, we get the following immediate consequence.

COROLLARY 2.4. *Let $j < h$ be integers, A, B sets in a commutative group and write $|A| = m$, $|A + jB| = \alpha m$. We have*

$$|hB| \leq \alpha^{h/j} m.$$

This is less general than Theorem 1.1, which will be proved in Section 8.

In the torsionfree case, using $|X + hB| \geq |X| + |hB| - 1$ instead (see exercises 4,8,9) we obtain the following result, which is stronger for α near to 1 (and gives the correct order of magnitude).

COROLLARY 2.5. *Let $j < h$ be integers, A, B sets in a torsionfree commutative group and write $|A| = m$, $|A + jB| = \alpha m$. We have*

$$|hB| \leq \left(\alpha^{h/j} - 1\right) m + 1.$$

EXERCISE 12. Let A, B be finite sets (in any commutative group), $|A| = n$, $|A+B| = \lambda n$. Show that there is a set T such that $|T| \le \lambda$ and $B \subset T + (A - A)$.

EXERCISE 13. Let A be a finite set (in any commutative group), $|A| = n$, $|2A| = \lambda n$. From Plünnecke's theorem we know that $|kA| \le \lambda^k n$. For fixed λ this is an exponential function of k. Find a bound of the form $f(k, \lambda)n$, where $f(k, \lambda)$ is, for fixed λ, a polynomial of k.

PREXERCISE. Let A be a finite set (in any commutative group). Prove that $|kA|$ is for $k > k_0$ actually equal to some polynomial of k (Hovanskii's theorem). (The polynomial and the value of k_0 depend on the set A.)

EXERCISE 14. Let $A \subset \mathbb{Z}$. Show that

$$k|(k+1)A| \ge (k+1)|kA| - 1.$$

The commutativity of the addition graph requires two assumptions: one is the commutativity of addition, the other is that the same set B is added repeatedly. Still, an application to different summands and noncommutative operation is possible; we will consider this in Sections 6 and 11.

Besides the complete addition graph we used above, a more general graph may be useful. Given three sets A, B, C we build on them the *restricted addition graph* as follows. The layers will be $V_0 = A$, $V_1 = (A + B) \setminus C$, $V_j = (A + jB) \setminus (C + (j-1)B)$ for $j > 1$. (We can omit this distinction by defining $0B = \{0\}$.) Again, there is an edge from an $x \in V_j$ to a $y \in V_{j+1}$ if $y = x + b$ with some $b \in B$. The case $C = \emptyset$ returns the complete addition graph. An important case is $C = A$, where in each stage we get the "new sums".

LEMMA 2.6. *The restricted addition graph is commutative.*

PROOF. Consider a typical path of length 2, $x \to y \to z$ with $x \in V_{j-1}$, $y \in V_j$, $z \in V_{j+1}$. This means $y = x + b$, $z = y + b'$ with $b, b' \in B$. We claim that $x \to x + b' \to x + b' + b = z$ is also a path in our graph. To see this we only need to check $x + b' \in V_j$, that is, $x + b' \in A + jB$ and $x + b' \notin C + (j-1)B$. The first follows from $x \in V_{j-1}$, and the negation of the second would imply $z = x + b' + b \in C + jB$, which would contradict $z \in V_{j+1}$.

We apply this substitution to a collection $x \to y \to z_i$ to find distinct y_i with $x \to y_i \to z_i$, and to a collection $x_i \to y \to z$ to find $x_i \to y_i \to z$; this is what we need to establish commutativity. □

By applying Plünnecke's theorem 2.1 to this graph we obtain the following.

THEOREM 2.7. *Let $j < h$ be integers, A, B, C sets in a commutative group and write $|A| = m$, $|(A + jB) \setminus (C + (j-1)B)| = \alpha m$. There is an $X \subset A$, $X \ne \emptyset$ such that*

$$|(X + hB) \setminus (C + (h-1)B)| \le \alpha^{h/j} |X|.$$

3. Magnification and disjoint paths

In this section we prove the following result.

THEOREM 3.1. *Let $\mathcal{G}$ be a commutative layered graph with layers $V_0, \dots, V_h$, $|V_0| = m$. If $\mu_h \ge 1$, then there are m (vertex)-disjoint paths from V_0 to V_h.*

The *outdegree* and *indegree* of a vertex x will be denoted by

$$d^+(x) = d^+(x, \mathcal{G}) = |\{y : x \to y\}|,$$
$$d^-(x) = d^-(x, \mathcal{G}) = |\{y : y \to x\}|.$$

LEMMA 3.2. *In a commutative graph if $x \to y$, then we have*

$$d^+(x) \geq d^+(y), \tag{3.1}$$

$$d^-(x) \leq d^-(y). \tag{3.2}$$

This is an immediate consequence of the definition of commutativity; we formulate it as a lemma to emphasize its importance.

DEFINITION 3.3. Given a graph $\mathcal{G} = (V, E)$ and two sets $X, Y \subset V$ of vertices, the *channel* between them is the graph $\overline{\mathcal{G}}(X, Y) = (\overline{V}, \overline{E})$ defined as follows. We take all directed paths starting in X and ending in Y, put all the vertices on these paths (including the endpoints) into $\overline{V}$ and connect two vertices if they are connected in $\mathcal{G}$. (It is easily seen that in a layered graph this is the same as putting all the vertices on the above mentioned paths into $\overline{E}$.)

LEMMA 3.4. *If $\mathcal{G}$ is commutative, so is every channel $\overline{\mathcal{G}}(X, Y)$.*

This is again an immediate consequence of the definition.

We see that the inequalities of Lemma 3.2 hold for every channel in a commutative graph, and this is the only property we will use.

EXERCISE 15. Suppose that a directed graph has the property that inequalities (3.1) and (3.2) hold for every channel in it. Is it necessarily commutative?

PROOF OF THEOREM 3.1. Let r be the maximal number of disjoint paths from V_0 to V_h. By Menger's theorem (see e. g. Ore [**36**] Ch. 12, or almost any book on graph theory) we know that there is a separating set S of cardinality r, that is, a set with the property that every path contains a vertex from S.

For a vertex $x \in V_i$ we say that i is its *index*, and denote it as $i = \operatorname{ind} x$.

From the separating sets of cardinality r we select one for which

$$\sum_{s \in S} \operatorname{ind} s \tag{3.3}$$

is minimal. We are going to show that

$$S \subset V_0 \cup V_h. \tag{3.4}$$

Let $\Omega_1, \dots, \Omega_r$ be r disjoint paths from V_0 to V_h. S has one element on each Ω_i, say s_i. Assume that (3.4) fails, and for some j, $1 \leq j \leq h-1$ we have

$$|S \cap V_j| = q > 0.$$

We may assume that

$$S \cap V_j = \{s_1, \dots, s_q\}.$$

For $1 \leq i \leq q$ let x_i be the predecessor and y_i the successor of s_i on Ω_i, so that

$$\Omega_i = (\dots, x_i, s_i, y_i, \dots).$$

The set

$$S' = \{x_1, \dots, x_q, t_{q+1}, \dots, t_r\}$$

cannot separate V_0 and V_h because of the minimality of the index sum (3.3). Consequently there is a path Γ from V_0 to V_h that avoids S'. It cannot avoid S, so it contains some vertex from $s_1, .., s_q$, say s_1. The predecessor x of s_1 on Γ is a vertex

$$x \notin \{x_1, \dots, x_q\}.$$

Write

$$M = \{s_1, \dots, s_q\},\ M^+ = \{y_1, \dots, y_q\},\ M^- = \{x_1, \dots, x_q\},$$
$$M' = M^- \cup \{x\},\ \mathcal{G}' = \overline{\mathcal{G}}(M', M^+).$$

We claim that the set of vertices of $\mathcal{G}'$ is $M' \cup M \cup M^+$. To see this suppose that there were a path Λ from x or from some x_i to some y_j that avoids M. In this case taking Γ or Ω_i from V_0 to x or x_i, then Λ to y_j, then Ω_j from y_j to V_h we would get a path from V_0 to V_h that avoids S, a contradiction.

Now we have the following chain of inequalities:

$$\begin{aligned}\sum_{i=1}^{q} d^+(x_i, \mathcal{G}') &\geq \sum_{i=1}^{q} d^+(s_i, \mathcal{G}') = \sum_{i=1}^{q} d^-(y_i, \mathcal{G}') \\ &\geq \sum_{i=1}^{q} d^-(s_i, \mathcal{G}') = \sum_{i=1}^{q} d^+(x_i, \mathcal{G}') + d^+(x, \mathcal{G}') \\ &> \sum_{i=1}^{q} d^+(s_i, \mathcal{G}'),\end{aligned}$$

a contradiction. Here the inequalities are applications of Lemma 3.2, the equalities express the fact that both sides are enumerations of the number of edges between M and M^+, and between M' and M, respectively. This contradiction proves (3.4). The separating property of S means that every upward path from $V_0 \setminus S$ must end in $V_h \cap S$. There are such paths (unless $S \supset V_0$, and in this case we are done), the assumption $\mu_h \geq 1$ means that the number of their possible endpoints is at least $|V_0 \setminus S|$, so we have

$$|V_h \cap S| \geq |V_0 \setminus S| = |V_0| - |V_0 \cap S|\,,$$

therefore

$$r = |S| = |V_h \cap S| + |V_0 \cap S| \geq |V_0|\,.$$

□

Corollary 3.5. *In a commutative graph if $\mu_h \geq 1$, then $\mu_j \geq 1$ for $1 \leq j \leq h$.*

Proof. Take a collection of m disjoint paths from V_0 to V_h. For any $X \subset V_0$ the paths that start from X cross V_j in $|X|$ different vertices that all belong to $\mathrm{im}(X, V_j)$. □

This is a particular case of Theorem 2.1 that will be used to deduce the general case in the next section.

4. Layered product

Definition 4.1. Let $\mathcal{G}' = (V', E')$ and $\mathcal{G}'' = (V'', E'')$ be h-layered graphs with layers V_i' and V_i'', resp. Their *layered product* is the h–layered graph on the layers $V_i = V_i' \times V_i''$, and two vertices $(x', x'') \in V_i$ and $(y', y'') \in V_{i+1}$ are connected if both $x' \to y'$ and $x'' \to y''$. This graph will be denoted by $\mathcal{G} = \mathcal{G}'\mathcal{G}''$. For repeated products with identical factors the usual power notation $\mathcal{G}^n$ will be used.

Observe that this is a proper subgraph of the usual product of these graphs.

LEMMA 4.2. *The layered product of commutative graphs is commutative as well.*

This is an immediate consequence of the definitions.

LEMMA 4.3. *Magnification ratios are multiplicative: if $\mathcal{G}, \mathcal{G}', \mathcal{G}''$ are h-layered graphs with magnification ratios μ_i, μ_i', μ_i'', resp., and $\mathcal{G} = \mathcal{G}'\mathcal{G}''$, then $\mu_i = \mu_i'\mu_i''$ for all i.*

PROOF. The inequality $\mu_i \leq \mu_i'\mu_i''$ is obvious: if μ_i' is attained at a subset $Z' \subset V_0$ and μ_i'' at $Z'' \subset V_0''$, then $Z = Z' \times Z'' \subset V_0$ gives the upper bound.

To prove the reverse inequality first consider a special case: $h = 1$, $\mathcal{G}''$ consists of two copies W_0, W_1 of a set W, and from a $w \in W_0$ there is a unique edge to the corresponding vertex in W_1, consequently $\mu_1'' = 1$.

Take a set

$$X \subset V_0 = V_0' \times W.$$

We have

$$X = \bigcup_{w \in W} X_w,$$

where X_w is the set of those elements of X whose second coordinate is w. We obtain

$$|\mathrm{im}(X, V_1)| = \sum |\mathrm{im}(X_w, V_1)| \geq \sum \mu_1' |X_w| = \mu_1' |X|$$

as desired.

Now consider the general case. We construct an auxiliary 3-layered graph $\mathcal{H}$ on the layers

$$U_0 = V_0' \times V_0'' = V_0,\ U_1 = V_j' \times V_0'',\ U_2 = V_j' \times V_j'' = V_j.$$

We connect an $(x', x'') \in U_0$ to $(y', x'') \in U_1$ (second coordinates equal) if there is a path from x' to y' in $\mathcal{G}'$, and we connect $(y', x'') \in U_1$ to $(y', y'') \in U_2$ (first coordinates equal) if there is a path from x'' to y'' in $\mathcal{G}''$. Clearly from (x', x'') to (y', y'') in $\mathcal{H}$ if and only if there is one in $\mathcal{G}$, so

$$\mu_2(\mathcal{H}) = \mu_j(\mathcal{G}).$$

The subgraphs $\mathcal{H}_1$, spanned by $U_0 \cup U_1$ and $\mathcal{H}_2$, spanned by $U_1 \cup U_2$ fall into the particular case treated above, which means

$$\mu_1(\mathcal{H}_1) = \mu(U_0, U_1) \geq \mu_j(\mathcal{G}'),\ \mu_1(\mathcal{H}_2) = \mu(U_1, U_2) \geq \mu_j(\mathcal{G}'').$$

Finally we have

$$\mu_2(\mathcal{H}) = \mu(U_0, U_2) \geq \mu(U_0, U_1)\mu(U_1, U_2) \geq \mu_j(\mathcal{G}')\mu_j(\mathcal{G}'')$$

by the previous inequality and this completes the proof of the reverse inequality $\mu_i \geq \mu_i'\mu_i''$. □

5. The independent addition graph

We define the *independent addition graph* $\mathcal{I}_{nh}$ as follows. Take a set B (say, of integers), $|B| = n$, such that all h-fold sums $b_1 + \cdots + b_h$, $b_i \in B$, are different, unless they are rearrangements of each other, and $A = \{0\}$, and build the addition graph on them. Since $|V_0| = 1$, the j-th magnification ratio of this graph is clearly

$$\mu_j(\mathcal{I}_{nh}) = |V_j| = |jB|\,.$$

EXERCISE 16. Calculate $|jB|$ as a function of j and n.

Since the number of formal j-fold sums is n^j and a sum occurs at most $j!$ times, we have

$$\frac{n^j}{j!} \le \mu_j(\mathcal{I}_{nh}) = |jB| \le n^j. \tag{5.1}$$

We shall also use the inverse of this graph. Here we have

$$\mu_h(\hat{\mathcal{I}}_{nh}) = |hB|^{-1} \ge n^{-h},$$

and for $j < h$

$$\mu_j(\hat{\mathcal{I}}_{nh}) \le \frac{|(h-j)B|}{|hB|} \le h! n^{-j}.$$

Exercise 17. Find the exact value of $\mu_j(\hat{\mathcal{I}}_{nh})$.

These graphs will be used in the proof of Plünnecke's theorem.

Proof of Theorem 2.1. We want to prove $\mu_h \le \mu_j^{h/j}$. We know that $\mu_j \ge 1$ whenever $\mu_h \ge 1$, and this settles the case $\mu_h = 1$.

Take now a graph $\mathcal{G}$ with $\mu_h < 1$. Consider the layered product $\mathcal{G}^* = \mathcal{G}^k \mathcal{I}_{nh}$. If we select k and n so that

$$\mu_h^k \frac{n^h}{h!} \ge 1,$$

then (with the natural notation) we have $\mu_h^* \ge 1$, hence $\mu_j^* \ge 1$, which then implies

$$\mu_j^k n^j \ge 1,$$

using the appropriate part of inequality (5.1).

To optimize this take

$$n = 1 + \left[\left(h! \mu_h^{-k} \right)^{1/h} \right] \le 2 h!^{1/h} \mu_h^{-k/h} = c_h \mu_h^{-k/h}.$$

The previous inequality gives

$$\mu_j \ge n^{-j/k} \ge c_h^{-1j/k} \mu_h^{j/k} \to \mu_h^{j/h}$$

as $k \to \infty$.

Finally assume that $\mu_h > 1$. Consider the layered product $\mathcal{G}^* = \mathcal{G}^k \hat{\mathcal{I}}_{nh}$. If we select k and n so that

$$\mu_h^k n^{-h} \ge 1,$$

then similarly we get $\mu_h^* \ge 1$ and hence $\mu_j^* \ge 1$, which then implies

$$\mu_j^k h! n^{-j} \ge 1,$$

using inequality (5.1) again.

Our choice of n is now

$$n = \left[\mu_h^{k/h} \right],$$

and the previous inequality gives

$$\mu_j \ge h!^{-1/k} n^{j/k} \to \mu_h^{j/h}$$

as $k \to \infty$. □

6. Different summands

An application to different summands is less straightforward, however, the case $j = 1$ of Theorem 2.3 can be extended in this way as follows [**47**].

THEOREM 6.1. *Let A, $B_1, \dots, B_h$ be sets in a commutative group G and write $|A| = m$, $|A + B_i| = \alpha_i m$. There is an $X \subset A$, $X \neq \emptyset$ such that*

$$|X + B_1 + \cdots + B_h| \le \alpha_1 \alpha_2 \dots \alpha_h |X|. \tag{6.1}$$

PROOF. Take auxiliary sets $T_1, \dots, T_h \subset G$ such that $|T_i| = n_i$ (which will be specified soon) and all the sums

$$y + t_1 + \cdots + t_h,\ y \in A + B_1 + \cdots + B_h,\ t_i \in T_i$$

are distinct. (This may be impossible in a finite group; in this case first embed it into an infinite one.) Now apply case $j = 1$ of Theorem 2.3 to the sets A and

$$B = \bigcup (B_i + T_i).$$

Observe that

$$|A + B| \le \sum |A + B_i + T_i| \le \sum |A + B_i|\, |T_i| = m \sum n_i \alpha_i.$$

We obtain the existence of a set $X \subset A$ such that

$$|X + hB| \le |X| \left(\sum n_i \alpha_i \right)^h.$$

On the other hand $X + hB \supset X + B_1 + \cdots + B_h + T_1 + \cdots + T_h$, consequently we have

$$|X + hB| \ge |X + B_1 + \cdots + B_h|\, n_1 \dots n_h.$$

A comparison of these inequalities gives

$$|X + B_1 + \cdots + B_h| \le \left(\sum n_i \alpha_i \right)^h (n_1 \dots n_h)^{-1} |X|\,. \tag{6.2}$$

To make this quotient small we put $n_i = n/\alpha_i$ with suitable n; since the numbers α_i are rational, this can be achieved with integers. Then inequality (6.2) turns into

$$|X + B_1 + \cdots + B_h| \le h^h \prod \alpha_i\, |X|\,, \tag{6.3}$$

which is worse than we claimed by a factor h^h.

To remove this factor we consider two 1-layered graphs. The first, say $\mathcal{G}$ is built on the sets A and $A + B_1 + \cdots + B_h$ in the natural way. The other, say $\mathcal{G}'$ is built similarly from the direct powers

$$A^k = A \times \cdots \times A, B_1^k, \dots, B_h^k$$

considered as sets in the k-th direct power of our initial group. Let μ and μ' be the magnification ratios of these graphs. The previous argument told us $\mu \le h^h \prod \alpha_i$, and the same, when applied to the sets A^k, B_j^k instead, gives

$$\mu' \le h^h \left(\prod \alpha_j \right)^k.$$

Now observe that $\mathcal{G}'$ is isomorphic to $\mathcal{G}^k$, so $\mu' = \mu^k$ by Lemma 4.3. The above inequality thus reduces to

$$\mu \le h^{h/k} \prod \alpha_j.$$

with an arbitrary k. As $k \to \infty$, we obtain (6.1). □

The case of general j can also be extended (a paper by Gyarmati, Matolcsi, Ruzsa in preparation [**18**])).

THEOREM 6.2. *Let $j < h$ be integers, and let A, $B_1, \dots, B_h$ be finite sets in a commutative group G. Let $K = \{1, 2, \dots, h\}$, and for any $I \subset K$ put*

$$B_I = \sum_{i \in I} B_i,$$

$$|A| = m, \quad |A + B_I| = \alpha_I m.$$

Write

$$\beta = \left(\prod_{L \subset K, |L| = j} \alpha_L \right)^{(j-1)!(h-j)!/(h-1)!}. \tag{6.4}$$

There exists an $X \subset A$, $X \neq \emptyset$ such that

$$|X + B_K| \leq \beta |X|. \tag{6.5}$$

7. Plünnecke's inequality with a large subset

We show an extension of Theorem 2.2 with a bound on the size of the selected subset.

THEOREM 7.1. *Let $j < h$ be integers, $\mathcal{G}$ a commutative layered graph on the layers $V_0, \dots, V_h$. Write $|V_0| = m$, $|V_j| = s$, $\gamma = h/j$. Let an integer k be given, $1 \leq k \leq m$. There is an $X \subset V_0$, $|X| \geq k$ such that*

(7.1)

$$|\operatorname{im}(X, V_h)| \leq \left(\frac{s}{m}\right)^\gamma + \left(\frac{s}{m-1}\right)^\gamma + \dots + \left(\frac{s}{m-k+1}\right)^\gamma + (|X| - k)\left(\frac{s}{m-k+1}\right)^\gamma.$$

PROOF. We use induction on k. The case $k = 1$ is Theorem 2.2.

Assume we know it for k; we prove it for $k+1$. The inductive assumption gives us a set X, $|X| \geq k$ with a bound on $|\operatorname{im}(X, V_h)|$ as given by (7.1). We want to find a set X' with $|X'| \geq k+1$ and

$$|\operatorname{im}(X', V_h)| \leq \left(\frac{s}{m}\right)^\gamma + \left(\frac{s}{m-1}\right)^\gamma + \dots + \left(\frac{s}{m-k}\right)^\gamma + (|X'| - k - 1)\left(\frac{s}{m-k}\right)^\gamma. \tag{7.2}$$

If $|X| \geq k+1$, we can put $X' = X$. If $|X| = k$, we apply Theorem 2.2 to the graph obtained from $\mathcal{G}$ by omitting the vertices in X. This yields a set $Y \subset V_0 \setminus X$ such that

$$|\operatorname{im}(Y, V_h)| \leq \left(\frac{s}{m-k}\right)^\gamma |Y|$$

and we put $X' = X \cup Y$. □

The following variant will be more comfortable for calculations.

THEOREM 7.2. *Let $j < h$ be integers, $\mathcal{G}$ a commutative layered graph on the layers $V_0, \dots, V_h$. Write $|V_0| = m$, $|V_j| = s$, $\gamma = h/j$. Let a real number t be given, $0 \leq t < m$. There is an $X \subset V_0$, $|X| > t$ such that*

$$|\operatorname{im}(X, V_h)| \leq \frac{s^\gamma}{\gamma}\left(\frac{1}{(m-t)^{\gamma-1}} - \frac{1}{m^{\gamma-1}}\right) + (|X| - t)\left(\frac{s}{m-t}\right)^\gamma. \tag{7.3}$$

PROOF. We apply Theorem 7.1 with $k = [t] + 1$. The right side of (7.3) can be written as $s^\gamma \int_0^{|X|} f(x)\,dx$, where $f(x) = (m-x)^{-\gamma}$ for $0 \le x \le t$, and $f(x) = (m-t)^{-\gamma}$ for $t < x \le |X|$. Since f is increasing, the integral is $\ge f(0) + f(1) + \cdots + f(|X|-1)$. This exceeds the right side of (7.1) by a termwise comparation. □

We state the consequences of this result for the complete and restricted addition graphs.

THEOREM 7.3. *Let $j < h$ be integers, A, B sets in a commutative group and write $|A| = m$, $|A + jB| = s$, $\gamma = h/j$. Let a real number t be given, $0 \le t < m$. There is an $X \subset A$, $|X| > t$ such that*

$$|X + hB| \le \frac{s^\gamma}{\gamma}\left(\frac{1}{(m-t)^{\gamma-1}} - \frac{1}{m^{\gamma-1}}\right) + (|X| - t)\left(\frac{s}{m-t}\right)^\gamma.$$

THEOREM 7.4. *Let $j < h$ be integers, A, B, C sets in a commutative group and write $|A| = m$, $|(A + jB) \setminus (C + (j-1)B)| = s$, $\gamma = h/j$. Let a real number t be given, $0 \le t < m$. There is an $X \subset A$, $|X| > t$ such that*

$$|(X + hB) \setminus (C + (h-1)B)| \le \frac{s^\gamma}{\gamma}\left(\frac{1}{(m-t)^{\gamma-1}} - \frac{1}{m^{\gamma-1}}\right) + (|X| - t)\left(\frac{s}{m-t}\right)^\gamma.$$

We state separately the case $j = 1$, $h = 2$ which will be applied in the sequel.

COROLLARY 7.5. *Let A, B sets in a commutative group and write $|A| = m$, $|A + iB| = s$. Let a real number t be given, $0 \le t < m$. There is an $X \subset A$, $|X| > t$ such that*

$$|X + 2B| \le \frac{s^2}{(m-t)^2}\left(|X| - \frac{t(t+m)}{2m}\right).$$

COROLLARY 7.6. *Let A, B, C be sets in a commutative group and write $|A| = m$, $|(A+B) \setminus C| = s$. Let a real number t be given, $0 \le t < m$. There is an $X \subset A$, $|X| > t$ such that*

$$|(X + 2B) \setminus (C + B)| \le \frac{s^2}{(m-t)^2}\left(|X| - \frac{t(t+m)}{2m}\right).$$

Theorem 6.1 can also be modified to yield large subsets.

THEOREM 7.7. *Let A, $B_1, \ldots, B_h$ be sets in a commutative group G and write $|A| = m$, $|A + B_i| = \alpha_i m$. Let a real number t be given, $0 \le t < m$. There is an $X \subset A$, $X \ne \emptyset$ such that*

$$|X + B_1 + \cdots + B_h| \le \alpha_1\alpha_2 \ldots \alpha_h m^h \left(\frac{1}{h}\left(\frac{1}{(m-t)^{h-1}} - \frac{1}{m^{h-1}}\right) + \frac{(|X| - t)}{(m-t)^{h-1}}\right). \tag{7.4}$$

The proof follows that of Theorem 7.1 with the difference that in the inductive step we apply Theorem 6.1 to the sets $A \setminus X$, $B_1, \ldots, B_h$. The available upper estimate for $|(A \setminus X) + B_i|$ is naturally $\alpha_i m$.

8. Sums and differences

With Plünnecke's method one can get various inequalities for cardinalities of sumsets, but it stops to work when differences are also involved (we shall give reasons why).

As far as I know the first inequality connecting sums and differences is due to Freiman and Pigaev [**11**]. They prove that

$$|A+A|^{3/4} \le |A-A| \le |A+A|^{4/3}. \tag{8.1}$$

We prove the following [**43**].

THEOREM 8.1. *Let A, Y, Z be finite sets in a (not necessarily commutative) group. We have*

$$|A||Y-Z| \le |A-Y||A-Z|. \tag{8.2}$$

PREXERCISE. Try to prove this inequality instead of reading the proof.

EXERCISE 18. Let

$$A = \{(x_1, \dots, x_d) \in \mathbb{Z}^d : x_i \ge 0, \sum x_i \le n\}.$$

a) Calculate $|A|$.
b) What are the elements of $A - A$?
c) What are the limits of $|A+A|/|A|$ and $|A-A|/|A|$ as $n \to \infty$ for fixed d?

PROOF. We will map the pairs $(a, x), a \in A, x \in Y - Z$ into $(A-Y) \times (A-Z)$ in an injective way.

List the elements of Y somehow, say $y_1, \dots, y_k$. Now given a pair (a, x), from all possible representations of x in the form $x = y - z$, $y \in Y$, $z \in Z$ select the one for which $y = y_i$ with minimal i, and map this pair into $(a - y, a - z)$. Take another pair (a', x') with representation $x' = y' - z'$. If we had

$$a - y = a' - y', \ a - z = a' - z',$$

then subtracting these equations (carefully in the noncommutative case!) we get $y - z = y' - z'$. Since both representations are minimal in the above sense, we conclude $y = y'$, then $z = z'$ and $a = a'$. □

Substituting $Y = -Y$ and $Z = -Z$, we obtain the following version.

$$|A||Y-Z| \le |Y+A||Z+A|. \tag{8.3}$$

Inequality (8.2) has the following interpretation (I cannot recall who made this observation). Define

$$\rho(X, Y) = \log \frac{|X-Y|}{\sqrt{|X|\,|Y|}}.$$

Then (8.2) can be written as

$$\rho(Y, Z) \le \rho(Y, A) + \rho(A, Z),$$

a triangle-inequality like property. ρ is also symmetric. A marked difference from distances is that $\rho(X, X)$ is typically positive.

EXERCISE 19. Show that $\rho(X, Y) \ge 0$, and find the cases of equality.

Substituting $Y = Z = -A$ in (8.2) we obtain the following inequality:

COROLLARY 8.2. *If $|A| = m$, $|2A| \le \alpha m$, then $|-A+A| \le \alpha^2 m$ and $|A-A| \le \alpha^2 m$.*

The second inequality above follows from the first via replacing A by $-A$ and observing that $2(-A) = -(2A)$.

We can also substitute $Y = Z = -2A$ to obtain the following inequality.

COROLLARY 8.3. *If* $|A| = m$, $|3A| \le \alpha m$, *then* $|-2A+2A| \le \alpha^2 m$.

The exponent 2 in Corollary 8.2 is best possible (though an improvement to something like $\alpha^2/(\log\alpha)^c$ is conceivable). This can be seen by considering the lattice points inside a d-dimensional simplex

$$\left\{(x_1,\dots,x_d)\in\mathbb{R}^d:\, x_i\ge 0,\, \sum x_i \le n\right\},$$

where $2^d \approx \alpha$. Denoting the volume of this simplex by v ($= n^d/d!$), the number of lattice points is about v, the size of the sumset is about the volume of this simplex dilated by 2, that is, $2^d v$, while the size of the difference set is about the volume of the difference set of this simplex, which is easily calculated to be $\binom{2d}{d}v$. Note that for a convex set the volume of the sumset is always 2^d times the original; the volume of the difference set varies and by a theorem of C. A. Rogers and G. G. Shephard [**41**], the simplex yields the maximum.

This example is analyzed in detail by Hennecart, Robert and Yudin [**22**]; they attribute the underlaying idea to Freiman and Pigaev's above mentioned paper [**11**]. See also A. Granville's survey [**15**], section 1.5.

One difference from the Plünnecke inequalities is the noncommutative nature of the above result. From Corollary 2.4 we obtain a similar implication: if $|A| = m$, $|A-A| \le \alpha m$, then $|2A| \le \alpha^2 m$. This fails in noncommutative groups, see Section 11.

Another difference is the following. To go to differences one would require the case "$h=-1$" of Theorem 2.3, which might be expected to sound as follows:

"if $|A| = n$, $|A+B| = \alpha n$, then there is a nonempty $X \subset A$ such that $|X-B| \le \alpha'|X|$, with α' depending only on α."

This is, however, false; we have the following results by Gyarmati, Hennecart, Ruzsa [**17**].

THEOREM 8.4. *Let* $\alpha > 2$. *Then for any* $c < \frac{\sqrt{2}\log 2}{\sqrt{3}}$ *and infinitely many* m, *there exist two sets* A *and* B *such that* $|A| = m$, $|A+B| \le \alpha m$ *and for any non-empty* $X \subset A$, *one has*

$$\frac{|X-B|}{|X|} \ge \exp\left(c\sqrt{(\log(\alpha/2))(\log m)(\log\log m)^{-1}}\right).$$

THEOREM 8.5. *Let* A *and* B *be non-empty and finite subset of some abelian group such that* $|A| = m$,$|A+B| \le \alpha m$. *Then there exists some non-empty subset* X *of* A *such that*

$$\frac{|X-B|}{|X|} \le \alpha \exp\left(2\sqrt{(\log\alpha)(\log m)}\right). \tag{8.4}$$

Inequality (8.2) together with Plünnecke's can be used to deduce the basic Theorem 1.1, which is sufficient for most of the applications ([**50**], Lemma 3.3) and which we repeat below.

THEOREM 8.6. *Let* A, B *be finite sets in a commutative group and write* $|A| = m$, $|A+B| = \alpha m$. *For arbitrary nonnegative integers* k, l *we have*

$$|kB - lB| \le \alpha^{k+l} m. \tag{8.5}$$

PROOF. By symmetry we may assume $k \leq l$. Assume also $k \geq 1$, since the case $k = 0$ is contained in Corollary 2.4. An application of Theorem 2.3 with $j = 1, h = k$ gives us a set $X \subset A$ such that

$$|X + kB| \leq \alpha^k |X| .$$

Another application with $j = k$, $h = l$ and X in the place of A gives a set $X' \subset X$ such that

$$|X' + lB| \leq \alpha^l |X'| .$$

Now apply Theorem 8.2 to the sets $-X'$, kB and lB to obtain

$$|X'| \, |kB - lB| \leq |X' + kB| \, |X' + lB| \leq \alpha^{k+l} |X'| \, |X| .$$

Now we divide by $|X'|$ and use $|X| \leq m$ to get inequality (8.5). □

The sum-sum analogue of Theorem 8.1 can be deduced from Plünnecke's inequality.

THEOREM 8.7. *In any commutative group we have*

$$|A||Y + Z| \leq |A + Y||A + Z|. \tag{8.6}$$

PROOF. Indeed, applying Theorem 6.1 we get a set $X \subset A$ such that

$$|X + Y + Z| \leq |X| \frac{|A + Y|}{|A|} \frac{|A + Z|}{|A|},$$

and to obtain (8.6) we just have to use $|X + Y + Z| \geq |Y + Z|$ and $|X| \leq |A|$. □

EXERCISE 20. Prove Freiman and Pigaev's inequality (8.1).

9. Double and triple sums

We present an inequality which sometimes nicely complements Plünnecke's.

THEOREM 9.1. *Let X, Y, Z be finite sets in a commutative group. We have*

$$|X + Y + Z|^2 \leq |X + Y||Y + Z||X + Z|. \tag{9.1}$$

This inequality may be extended to the noncommutative case as follows.

THEOREM 9.2. *Let X, Y, Z be finite sets in a not necessarily commutative group. We have*

$$|X + Y + Z|^2 \leq |X + Y||Y + Z| \max_{y \in Y} |X + y + Z|. \tag{9.2}$$

PROOF. We use induction on $|Y|$. For $|Y| = 1$ (9.2) reduces to the obvious inequality

$$|X + y + Z| \leq |X||Z|.$$

Assume now we know (9.2) for smaller sets. Fix y as the element of Y which maximizes $|X + y + Z|$. Write

$$|X + y + Z| = m, Y \setminus \{y\} = Y',$$

$$|(X + Y + Z) \setminus (X + Y' + Z)| = a, |(X + Y) \setminus (X + Y')| = b, |(Y + Z) \setminus (Y' + Z)| = c.$$

With these notations (9.2) can be rewritten as

$$(|X + Y' + Z| + a)^2 \leq m(|X + Y'| + b)(|Y' + Z| + c). \tag{9.3}$$

We shall obtain (9.3) as the sum of the following three inequalities:

$$|X + Y' + Z|^2 \leq m|X + Y'||Y' + Z|, \tag{9.4}$$

$$2a|X+Y'+Z| \le m(c|X+Y'|+b|Y'+Z|), \tag{9.5}$$

$$a^2 \le mbc. \tag{9.6}$$

Of these inequalities (9.4) follows from the induction hypothesis.

Clearly every element of $(X+Y+Z)\setminus(X+Y'+Z)$ is of the form $x+y+z$ with $x\in X$, $z\in Z$, hence $a\le m$. We can map this set into the Cartesian product of $(X+Y)\setminus(X+Y')$ and $(Y+Z)\setminus(Y'+Z)$ by mapping a typical element $x+y+z$ into the pair $(x+y, y+z)$. This pair determines $x+y+z$ uniquely and clearly $x+y\notin X+Y'$ as otherwise we would have $x+y+z\in X+Y'+Z$; similarly $y+z\notin Y'+Z$. This mapping shows $a\le bc$. The product of these inequalities gives (9.6).

By multiplying inequalities (9.4) and (9.6) and taking the square root we obtain

$$a|X+Y'+Z| \le m\sqrt{bc|X+Y'||Y'+Z|};$$

(9.5) now follows from the arithmetic-geometric mean inequality. □

In the commutative case this inequality can be extended to more than 3 sets as follows (Gyarmati, Matolcsi, Ruzsa [**19**]).

THEOREM 9.3. *Let $A_1,\dots,A_k$ be finite, nonempty sets in an arbitrary commutative semigroup. Put*

$$S = A_1+\dots+A_k,$$
$$S_i = A_1+\dots+A_{i-1}+A_{i+1}+\dots+A_k.$$

We have

$$|S| \le \left(\prod_{i=1}^{k}|S_i|\right)^{\frac{1}{k-1}}. \tag{9.7}$$

Curiously, one of the arguments relies on invertibility, the other on commutativity, so we do not have any result for noncommutative semigroups. Neither could we extend the above noncommutative argument for more than three summands, and hence the following question remains open.

PROBLEM 9.4. Let $A_1,\dots,A_k$ be finite, nonempty sets in an arbitrary noncommutative group. Put

$$S = A_1+\dots+A_k,$$
$$n_i = \max_{a\in A_i}|A_1+\dots+A_{i-1}+a+A_{i+1}+\dots+A_k|.$$

Is it true that

$$|S| \le \left(\prod_{i=1}^{k} n_i\right)^{\frac{1}{k-1}} ? \tag{9.8}$$

We finish this section by a meditation on the sizes of $2A$ and $3A$.

Write $|A|=m$, $|2A|=n$. Corollary 2.4 implies $|3A|\le n^3/m^2$, and Theorem 9.1 implies $|3A|\le n^{3/2}$. The first is better for $n\le m^{4/3}$, the second for larger values. The two together describe the maximal possible value of $|3A|$ up to a constant.

THEOREM 9.5. *Let m,n be positive integers satisfying $m\le n\le m^2$. There is a set A of integers such that $|A|\asymp m$, $|2A|\asymp n$ and*

$$|3A| \asymp \min\left(n^3/m^2, n^{3/2}\right).$$

PROOF. We construct A in $\mathbb{Z}^3$. Take two integers k, l such that $k \le l \le k^3$ and put

$$A_1 = \{(x, y, z) : \ 0 \le x, y, z < k\},$$

$$A_2 = \{(x, 0, 0), (0, x, 0), (0, 0, x) : \ 0 \le x < l\},$$

and $A = A_1 \cup A_2$.

We have $m = |A| = k^3 + 3(l - k) \asymp k^3$, so the proper choice is $k \sim m^{1/3}$. Further $2A = 2A_1 \cup (A_1 + A_2) \cup 2A_2$. The cardinality of the parts is of order k^3, $k^2 l$ and l^2, respectively. The first is always smaller than the second, hence

$$n = |2A| \asymp \max(k^2 l, l^2);$$

the threshold of behaviour is at $l = k^2$. Hence the proper choice of l is

$$l \sim \min(\sqrt{n}, n/m^{2/3})$$

and the claim follows from the fact that $|3A| \ge |3A_2| \ge l^3$. □

10. $A + B$ and $A + 2B$

In this section we consider the following problem. Let $|A| = m$, $|A+B| = \alpha m$. How large can $|A + 2B|$ be? In the case $B = A$ the answer was given at the end of the last section. A similar bound can be found by Plünnecke's method if A and B are about the same size. Without any assumption on B, however, the situation changes.

An application of Theorem 9.1 immediately yields

$$|A + 2B| \le |A + B|\sqrt{|2B|}; \tag{10.1}$$

this inequality was already proved differently in [**56**], Theorem 7.2. To estimate $|2B|$ we can use Corollary 2.4 to obtain $|2B| \le \alpha^2 m$; combined with (10.1) we get

$$|A + 2B| \le \alpha^2 m^{3/2}.$$

In [**56**], Theorem 7.1 examples are given (for every rational α and infinitely many m) such that

$$|A + 2B| \ge \left(\frac{\alpha - 1}{4}\right)^2 m^{3/2}. \tag{10.2}$$

These results describe the order of magnitude for fixed $\alpha > 1$ unless α is near to 1.

We now explore what happens for small values of α. In the extremal case $\alpha = 1$ clearly also $A + 2B = m$. The transition is somewhat less clear.

THEOREM 10.1. *Let A, B be finite sets in a commutative group G, $|A| = m$, $|A + B| = \alpha m$, $1 < \alpha \le 2$. We have*

$$|A + 2B| \le \alpha m + \frac{3}{2}(\alpha - 1)m\sqrt{|2B|}, \tag{10.3}$$

consequently

$$|A + 2B| \le \alpha m + 3(\alpha - 1)m^{3/2}; \tag{10.4}$$

if G is torsionfree, then

$$|A + 2B| \le \alpha m + 3(\alpha - 1)^{3/2} m^{3/2}; \tag{10.5}$$

PROOF. We apply Corollary 7.6 with the choice $C = A + b$, where b is an arbitrary element of B. The s in the hypothesis will be

$$s = |(A + B) \setminus (A + b)| = (\alpha - 1)m,$$

and we obtain (for every $0 \le t < m$) the existence of an $X \subset A$, $|X| > t$ such that

$$|(X + 2B) \setminus (C + B)| \le \frac{s^2}{(m-t)^2}\left(|X| - \frac{t(t+m)}{2m}\right).$$

Since $|C + B| = |A + B| = \alpha m$, this implies

$$|(X + 2B)| \le \alpha m + \frac{s^2}{(m-t)^2}\left(|X| - \frac{t(t+m)}{2m}\right).$$

For $A \setminus X$ we use an obvious estimate:

$$|(A \setminus X) + 2B| \le |A \setminus X||2B| = (m - |X|)|2B|,$$

and sum the last two inequalities to get

$$|(A + 2B)| \le \alpha m + \left(\frac{s^2}{(m-t)^2} - |2B|\right)|X| + m|2B| - \frac{s^2}{(m-t)^2}\frac{t(t+m)}{2m}. \tag{10.6}$$

We choose t so that the coefficient of $|X|$ vanishes, that is,

$$\frac{s^2}{(m-t)^2} = |2B|. \tag{10.7}$$

Such a t exists in the interval $(0, m)$ as long as $|2B| \ge s^2/m^2 = (\alpha-1)^2$, which certainly holds under our assumption $\alpha \le 2$. (We do not really need this restriction; however, for $\alpha > 2$ this estimate is weaker than (10.1), due to the factor 3/2.) With this choice (10.6) becomes

$$|(A + 2B)| \le \alpha m + |2B|\left(m - \frac{t(t+m)}{2m}\right) = \alpha m + |2B|\frac{(m-t)(2m+t)}{2m}. \tag{10.8}$$

We estimate $2m + t$ by $3m$, and we express $m - t$ by (10.7):

$$m - t = \frac{s}{\sqrt{|2B|}} = \frac{(\alpha-1)m}{\sqrt{|2B|}}.$$

After these substitutions (10.8) becomes (10.3).

To deduce (10.4) we use Corollary 2.4 and $\alpha \le 2$.

To deduce (10.5) we use Corollary 2.5: in a torsionfree group

$$|2B| \le 1 + (\alpha^2 - 1)m \le 4(\alpha - 1)m,$$

since $\alpha = |A + B|/m \ge 1 + 1/m$, and put this into (10.3). □

We remark that the summand αm in these estimates can actually be the main term, as α may be as small as $1 + O(1/m)$. In the general estimate (10.4) the threshold is $1 + O(m^{-1/2})$, in the torsionfree estimate (10.5) it is $1 + O(m^{-1/3})$.

Still there is a gap between the exponent 2 of $\alpha - 1$ in the example (10.2) and 3/2 in the upper estimate (10.5). We now show by an example that the exponent 1 of $\alpha - 1$ for general groups in (10.4) is exact. Take a group G which has two k-element subgroups H_1, H_2 such that $H_1 \cap H_2 = \{0\}$. Write $H = H_1 + H_2$ and let

$$A = H \cup \{a_1, \dots, a_t\}, \quad B = H_1 \cup H_2,$$

where $a_1, \dots, a_t$ lie in different nonzero cosets of H. Observe that $2B = H$. We have

$$m = |A| = k^2 + t,$$

$$\alpha m = |A + B| = k^2 + t(2k - 1),$$

$$\alpha - 1 = \frac{2t(k-1)}{m},$$

$$|A + 2B| = (t + 1)k^2 = \alpha m + (\alpha - 1)(k - 1)m/2.$$

Since $k - 1 \sim \sqrt{m}$ as long as $t = o(k^2)$ (and in the interesting case $t = O(k)$), the only difference from the upper estimate (10.4) is a factor of 6.

11. On the noncommutative case

Our attention was focused on commutative groups, with special emphasis on itegers. At several places, namely at Theorem 8.1 and Theorem 9.2, we mentioned the possibility of a noncommutative extension. We now explore the limits of this extension.

First we collect some examples that show how certain attempts of extension fail. These examples use a free group, which is "very noncommutative"; it is possible that for groups "nearer" to commutative ones in some sense some results can be extended.

First recall some results that did not require commutativity. Thoerem 8.1 told us

$$|A||Y - Z| \le |A - Y||A - Z|. \tag{11.1}$$

This had the following consequences (Corollary 8.2): if $|A| = m$, $|2A| \le \alpha m$, then $|-A + A| \le \alpha^2 m$ and $|A - A| \le \alpha^2 m$.

We first show that the two cases in the above corollary are not superfluous, in a noncommutative group $|-A + A|$ and $|A - A|$ can be very different (of course, without the assumption on $2A$).

Indeed, take a free group generated by the elements a, b and put

$$A = \{ia + b : 1 \le i \le m\} \cup \{ia : 1 \le i \le m\}.$$

Then $|A| = 2m$ and

$$-A = \{-b - ja : 1 \le j \le m\} \cup \{-ja : 1 \le j \le m\}.$$

Here $A - A$ contains the $2m^2$ different elements $ia \pm b - ja$, while

$$-A + A = \{(i - j)a\} \cup \{(i - j)a + b\} \cup \{-b + (i - j)a\} \cup \{-b + (i - j)a + b\},$$

altogether $4m$ elements.

So if the sumset is small, both difference sets are small without commutativity. In the commutative case from Corollary 2.4 we obtain a similar implication: if $|A| = m$, $|A - A| \le \alpha m$, then $|2A| \le \alpha^2 m$. This also fails in noncommutative groups. As above, take a free group with generators a, b and put

$$A = \{ia + b : 1 \le i \le m\}.$$

Then both difference sets $A - A$ and $-A + A$ have $2m - 1$ elements, while $|2A| = m^2$.

Between double and triple sums we had the following inequality without commutativity (Theorem 9.2).

$$|X + Y + Z|^2 \le |X + Y||Y + Z| \max_{y \in Y} |X + y + Z|. \tag{11.2}$$

We show by an example that the maximum cannot be omitted and cannot even be replaced by an average, even in the case of identical sets. As before, take a free group with generators a, b and put

$$X = Y = Z = \{a, 2a, \dots, na, b\}.$$

We have $|X| = n + 1$, $|2X| = 4n$ and $|3X| > n^2$ since all the elements $ia + b + ja$, $1 \le i, j \le n$ are distinct. From the $n + 1$ sets $X + y + X$, $y \in X$ only one is of size n^2, namely the one with $y = b$, all the others have $O(n)$ elements.

For a constrast, by applying Corollary 2.4, with similar values of $|X|$ and $|2X|$ in a commutative group we would have $|3X| \le 4^3 n$.

These examples suggest that commutativity is not only an assumption heavily used in Plünnecke's method, but a typical result will fail without it. The last example suggest a possible noncommutative replacement.

Problem 11.1 (A noncommutative Plünnecke?). Theorems 9.1 and 9.2 suggest a way to find noncommutative analogues of inequalities that for commutative groups were proved by Plünnecke's method. I formulate the simplest possible of them. Let A, B be finite sets in a noncommutative group, and define α by

$$\max_{b \in B} |A + b + B| = \alpha |A|.$$

Must there exist a nonempty $X \subset A$ such that

$$|X + 2B| \le \alpha' |X|$$

with an α' depending only on α?

I rather expect a negative answer.

However, Plünnecke's method can be modified to handle some noncommutative situations. We give a simple example of this.

Theorem 11.2. *Let A, B_1, B_2 be sets in a (typically noncommutative group) G and write $|A| = m$, $|B_1 + A| = \alpha_1 m$, $|A + B_2| = \alpha_2 m$. There is an $X \subset A$, $X \ne \emptyset$ such that*

$$|B_1 + X + B_2| \le \alpha_1 \alpha_2 |X|. \tag{11.3}$$

Proof. We take 4 copies of G and build a 2-layered graph on them. V_0 contains the set A in one copy, V_1 contains the sets $B_1 + A$ and $A + B_2$ in different copies, and V_2 contains $B_1 + A + B_2$; edges are drawn in the natural way.

We claim that this graph is commutative. Indeed, take vertices such that $x \to y \to z_i$, $i = 1, \dots, k$. The edge $x \to y$ can go either to $B_1 + A$ or $A + B_2$; assume the first, the other is similar. Then $x \in A$, $y = b + x$ with some $b \in B_1$ and $z_i = b + x + c_i$ with $c_i \in B_2$. Then with $y_i = x + c_i$ we have $x \to y_i \to z_i$; observe that these replacing vertices are in the other half of V_1 than the original one.

The other side of commutativity goes similarly. If we have $x_i \to y \to z$, and $y \in B_1 + A$, then there are elements $b_i \in B_1$ such that $b_i + x_i = y$, and $c \in B_2$ such that $z = y + c$. The replacing edges are again $x_i \to x_i + c \to b_i + x_i + c = y + c = z$ in the other half of the graph.

Note that this was based on a special kind of commutativity: adding an element from the left and adding another from the right commute — this property bears the name "associativity".

Applying Plünnecke's graph theorem 2.1 to this graph we obtain a set $X \subset A$ such that

$$|B_1 + X + B_2| \leq (\alpha_1 + \alpha_2)^2 |X| . \tag{11.4}$$

If $\alpha_1 = \alpha_2 = \alpha$, this is $4\alpha^2$ rather than the α^2 claimed in (11.3), if they are rather different, it can be much worse. We can improve the situation by embedding G into a larger group $G' = G \times H_1 \times H_2$, where the H_i are cyclic groups, $|H_i| = n_i$, and we identify G with $G \times \{0\} \times \{0\}$. In G' we consider the sets $A' = A$, $B_i' = B_i \times H_i$. We have $\alpha_i' = \alpha_i n_i$, and an application of (11.4) yields an $X \subset A$ such that

$$|B_1' + X + B_2'| = n_1 n_2 |B_1 + X + B_2| \leq (\alpha_1 n_1 + \alpha_2 n_2)^2 |X| .$$

If we select the n_i so that $\alpha_1 n_1 = \alpha_2 n_2$, this gives

$$|B_1 + X + B_2| \leq 4\alpha_1 \alpha_2 |X| \tag{11.5}$$

in the general case.

We can remove the factor 4 like in the previous proof. We take the 1-layered graph built on the layers A and $B_1 + A + B_2$. We are interested in the magnification ratio μ of this graph. We take the similar graph made from the sets A^k, B_i^k. This graph is the same as the k-th power of the previous graph, thus its magnification ratio is μ^k. An application of (11.5) gives $\mu^k \leq 4(\alpha_1 \alpha_2)^k$; taking k -th roots and making $k \to \infty$ we obtain (11.3). □

We mention without proof how this result can be generalized to several summands, with an extra condition.

Definition 11.3. A collection of sets $B_1, \ldots, B_k$ in a (noncommutative) group is *exocommutative*, if for all $x \in B_i$, $y \in B_j$ with $i \neq j$ we have $x + y = y + x$.

Theorem 11.4. *Let $A, B_1, B_2, \ldots, B_k, C_1, C_2, \ldots, C_l$ be sets in a (typically noncommutative group) G and write $|A| = m$, $|B_i + A| = \alpha_i m$, $i = 1, \ldots, h$, $|A + C_j| = \beta_j m$ $j = 1, \ldots, l$. Assume that both $B_1, \ldots, B_k$ and $C_1, \ldots, C_l$ are exocommutative. Then there is an $X \subset A$, $X \neq \emptyset$ such that*

$$|B_1 + \ldots + B_k + X + C_i + \ldots + C_l| \leq \alpha_1 \ldots \alpha_k \beta_1 \ldots \beta_l |X|. \tag{11.6}$$

The moral seems to be that it is hopeless to undestand sets such that $|2A| \leq \alpha |A|$ in general groups. On the other hand, if we start with an assumption on a threefold sum, say $|A| = m$, $|3A| \leq gam$, then an iterated application of Theorem 8.1 gives estimates for arbitary sum-difference combinations. For instance, putting $Y = Z = -2A$ we get $|-2A + 2A| \leq \alpha^2 m$, then with $Y = Z = A - 2A$ and putting $-A$ into the place of A we get an estimate for a sixfold sumset and so on. It is the step from 2 to 3 which fails in lack of commutativity.

Finally we add that a weaker conclusion can be drawn from the assumption $|2A| \leq \alpha m$, namely, that there is an $A' \subset A$, $|A'| > (1 - \varepsilon)m$ such that $3A'$ and hence each kA' is small.

To this end we apply an argument like in Section 7. Theorem 11.2 above assures the existence of a nonempty set X such that, if $|L + A| \leq \alpha m$ and $|A + R| \leq \beta m$ then $|L + X + R| \leq \alpha\beta |X|$.

Choose an $\epsilon > 0$, let $m = |A|$ and define $X_1 = X$. If $|X_1| > (1 - \epsilon)|A|$ we are done. If this is not the case we apply Theorem 11.2 on $A_1 = A \setminus X_1$, with this procedure we obtain an $X_2 \in A_1$ with a similar property. If $X_1 \cup X_2$ is still not large enough, we

continue with the procedure till we get $X' = X_1 \cup \ldots \cup X_h$ for some h and such that $|X'| > (1-\epsilon)m$. To bound $|L + X' + R|$ we need to estimate

$$\frac{|L + A_i|}{|A_i|}$$

for all $1 \leq i \leq h$. Such a bound is $\alpha|A|/|A_h| \leq \alpha/\epsilon$, since $|A_h| \geq \epsilon|A|$ and $|L + A_i| \leq |L + A|$. Similarly

$$\frac{|A_i + R|}{|A_i|} \leq \frac{\beta}{\epsilon}|A|,$$

for all i. Hence, by adding all the pieces X_i, we obtain

$$|L + A' + R| \leq \frac{\alpha\beta}{\epsilon^2}|A'|.$$

We had pay an ϵ^2 price to get $|A'| > (1-\varepsilon)\,|A|$; this can be somewhat improved with a more involved argument like in Section ref g.

CHAPTER 2

Structure of sets with few sums

1. Introduction

We want to describe sets that have few sums. If $|A| = m$, then clearly $|A + A| \geq m$ in every group (with equality for cosets), which can be improved to $2m - 1$ for sets of integers (or torsionfree groups in general). What can we say if we know that $|A + A| \leq \alpha m$, where α is constant or grows slowly as $n \to \infty$? That is, we are looking for statements of the form

$$|A| = m,\ |A + A| \leq \alpha m \Longrightarrow (\dots).$$

Such a condition $(\dots)$ is *adequate*, if this implication can be reversed to some degree, that is, there is an implication in the other direction

$$(\dots) \Longrightarrow |A + A| \leq \alpha' m,$$

with $\alpha' = \alpha'(\alpha)$ depending only on α and not on m or other properties of the set.

Between such results we can distinguish on two grounds. First, the smaller the value of α', the better the description; next, subjectively, the more we learn on the structure of the set the happier we are.

As an example consider the following implications (see chapter 1, Section 8):

$$|A| = m,\ |A + A| \leq \alpha m \Longrightarrow |A - A| \leq \alpha^2 m$$

and

$$|A| = m,\ |A - A| \leq \alpha m \Longrightarrow |A + A| \leq \alpha^2 m.$$

If we combine the two we get that

$$|A + A| \leq \alpha m \Longrightarrow |A - A| \leq \alpha^2 m \Longrightarrow |A + A| \leq \alpha^4 m,\ \alpha' = \alpha^4,$$

so this is an adequate description with a very good value of α', but it tells little about the structure of A and it is not surprising. Indeed,

$$a + b = c + d \Longleftrightarrow a - c = d - b,$$

so a coincidence between sums corresponds to a coincidence between differences. In particular, this shows that $|A + A|$ attains its maximal value $m(m + 1)/2$ exactly when $|A - A|$ attains its maximal value $m(m-1)+1$. (Such sets, with no nontrivial coincidence between sums or differences, are often called *Sidon sets*.)

There is a similar connection between minimal values of these quantities. For sets of integers the minimal value of both $|A + A|$ and $|A - A|$ is $2m - 1$, and equality occurs only for arithmetic progressions.

Still, the connection here is less obvious than it looks. We illustrate this by the case of near-maximal values. Suppose that $|A + A| \geq \kappa m^2$; does it follow that $|A - A| \geq \kappa' m^2$ with some κ' depending on κ? The answer is negative in a rather strong way: $|A + A| > m^2/2 - m^{2-\delta}$ and $|A - A| < m^{2-\delta}$ can happen with some constant $\delta > 0$. Similarly $|A - A| > m^2/2 - m^{2-\delta}$ and $|A + A| < m^{2-\delta}$ is also possible [**52**].

A set of integers with a minimal sumset ($|A + A| = 2m - 1$) is necessarily an arithmetic progression. This easy result exhibits some stability. A set with a nearly minimal sumset is almost an arithmetic progression, as the following result shows.

THEOREM 1.1 (G. Freiman[**10**]). *If $A \subset \mathbb{N}$, $|A| = m$, $|A + A| \le 3m - 4$, then A is contained in an arithmetic progression of length $\le |A + A| - m + 1 \le 2m - 3$.*

(Proof in the next chapter.)

Beyond $3m$, however, a single arithmetic progression is insufficient, as the following example shows. Take

$$A = \{1, \dots, m/2\} \cup \{t + 1, \dots, t + m/2\},$$

. . . .

we have $|A + A| = 3m - 3$, and A cannot be covered by a progression shorter than $t + m/2$. The reason is that this set has a hidden two-dimensional structure:

. . . . . .

. . . . . .

These sets are not isomorphic algebraically, but they behave analogously regarding the coincidence of sums. To describe such sets we need multidimensional, or generalized arithmetic progressions.

DEFINITION 1.2. Let $q_1, \dots, q_d$ and a be elements of an arbitrary commutative group, $l_1, \dots, l_d$ positive integers. A d-dimensional *generalized arithmetic progression* is a set of the form

$$P = P(q_1, \dots, q_d; l_1, \dots, l_d; a) = \{a + x_1 q_1 + \cdots + x_d q_d : 0 \le x_i \le l_i\} \tag{1.1}$$

(a projection of a cube). More exactly, we think of it as a set together with a fixed representation in the form (1.1); this representation is in general not unique. We call d the *dimension* of P, and by its *size* we mean the quantity

$$\|P\| = \prod_{i=1}^{d} (l_i + 1),$$

which is the same as the number of elements if all sums in (1.1) are distinct. In this case we say that P is *proper*.

EXERCISE 21. If P is a d-dimensional progression, then

$$|2P| < 2^d |P| \le 2^d \|P\|$$

.

The principal result sounds as follows.

THEOREM 1.3 (G. Freiman[**10**]). *If $A \subset \mathbb{Z}$, $|A| = n$, $|A + A| \le \alpha n$, then A is contained in a generalized arithmetic progression of dimension $\le d(\alpha)$ and size $\le s(\alpha)n$.*

This is an adequate description with the simplest possible structure: if $A \subset P$, then

$$|A + A| \le |P + P| < 2^d \|P\| \le 2^d s n,$$

$$\alpha' = 2^{d(\alpha)} s(\alpha).$$

More generally we have

$$|kA| \leq k^d \|P\|$$

. This shows that this dimension is closely connected with the rate of growth of $|kA|$ as a function of k.

For a comprehensive account of this theory up to 1996 see Nathanson's book [**33**].

Three basic questions arise here:

1) to find good bounds for $d(\alpha)$, $s(\alpha)$;

2) is this the "real" form?

3) how to extend this from $\mathbb{Z}$ to other groups.

Bounds: due to works by the author [**50, 53**], Y. Bilu [**2**], M. C. Chang [**5**] we know that $d < \alpha$ (best possible) and $s < e^{\alpha^c}$. It is also known that a bound for s must be $\gg 2^\alpha$; probably the proper order is $e^{c\alpha}$.

The real form: probably a flexible form (several covering sets, projections of lattice points in more general convex bodies) would give better bounds for α'.

Other groups. For sets situated in $\mathbb{Z}^m$ or in general commutative torsionfree groups verbatim the same result holds (and later we shall formulate and prove it in this setting).

In groups with torsion a new phenomenon arises, namely any coset has $|A+A| = |A|$. For groups with a strong torsion property this alone suffices to characterize sets with small sumsets.

Recall that the *exponent* of a group G is the smallest positive integer r such that $rg = 0$ for every $g \in G$.

THEOREM 1.4. *Let G be a commutative group of exponent r, $A \subset G$, $|A| = m$, $|A + A| \leq \alpha m$. A is contained in a coset of a subgroup of size $\leq \alpha^2 r^{\alpha^4} m$.*

We shall start (in the next section) with the proof of this theorem, which is simple and highlights some aspects of the case of integers.

General commutative groups

In a general commutative group, a set with a small sumset can be covered by a combination of the two mentioned structures, cosets and generalized arithmetic progressions.

THEOREM 1.5 (Green-Ruzsa [**16**]). *Let G be a commutative group, $A \subset G$, $|A| = m$, $|A+A| \leq \alpha m$. A is contained in a set of the form $H+P$, where H is a subgroup, P is a generalized arithmetic progression, the dimension of P is $\leq d(\alpha)$ and $|H||P| \leq s(\alpha)m$.*

For the quantities we have the following bounds: $d(\alpha) \ll \alpha^c$, $s(\alpha) \ll e^{\alpha^c}$.

Noncommutative groups.

For general groups, I do not even have a decent conjecture. There is a structure theorem for $SL_2(\mathbb{R})$ (Elekes-Király[**6**]). Roughly speaking, it asserts that a set with a small sumset is contained in a few cosets of a commutative subgroup, and within a coset we have a generalized arithmetic progression structure.

2. Torsion groups

In this section we prove Theorem 1.4, in a superficially more general form.

THEOREM 2.1. *Let $r \geq 2$ be an integer, and let G be a commutative group of exponent r. Let $A \subset G$ be a finite set, $|A| = m$. If there is another set $A' \subset G$ such that $|A'| = m$ and $|A + A'| \leq \alpha m$ (in particular, if $|A + A| \leq \alpha m$ or $|A - A| \leq \alpha m$), then A is contained in a subgroup H of G such that*

$$|H| \leq f(r, \alpha)m,$$

where

$$f(r, \alpha) = \alpha^2 r^{\alpha^4}.$$

PROOF. Let $b_1, b_2, \dots, b_k$ be a maximal collection of elements such that $b_i \in 2A - A$ and the sets $b_i - A$ are all disjoint. We have

$$b_i - A \subset 2A - 2A,$$

hence

$$\Big|\bigcup(b_i - A)\Big| = km \leq |2A - 2A| \leq \alpha^4 m$$

(the last inequality follows from Theorem 1.1 of Chapter 1). This implies $k \leq \alpha^4$.

Take an arbitrary $x \in 2A - A$. Since the collection $b_1, \dots, b_k$ was maximal, there must be an i such that

$$(x - A) \cap (b_i - A) \neq \emptyset,$$

that is, $x - a_1 = b_i - a_2$ with some $a_1, a_2 \in A$, which means

$$x = b_i + a_1 - a_2 \in b_i + (A - A).$$

Hence

$$2A - A \subset \bigcup\big(b_i + (A - A)\big) = B + A - A, \tag{2.1}$$

where $B = \{b_1, \dots, b_k\}$.

Now we prove

$$jA - A \subset (j - 1)B + A - A \quad (j \geq 2) \tag{2.2}$$

by induction on j. By (2.1) , this holds for $j = 2$. Now we have

$$\begin{aligned}(j + 1)A - A &= (2A - A) + (j - 1)A \\ &\subset B + A - A + (j - 1)A \text{ by (2.1)} \\ &= B + (jA - A) \\ &\subset B + (j - 1)B + A - A \\ &= jB + A - A,\end{aligned}$$

which provides the inductive step.

Let H and I be the subgroups generated by A and B, respectively. By (2.2) we have

$$jA - A \subset I + (A - A) \tag{2.3}$$

for every j. We have also

$$\bigcup(jA - A) = H, \tag{2.4}$$

which easily follows from the fact that the order of the elements of G is bounded. (2.3) and (2.4) imply that

$$H \subset I + (A - A).$$

Since I is generated by k elements of order $\leq r$ each, we have

$$|I| \leq r^k \leq r^{\alpha^4},$$

consequently

$$|H| \leq |I||A - A| \leq \alpha^2 r^{\alpha^4} m$$

(the estimate for $|A - A|$ follows again from Theorem 1.1, Ch. 1). □

REMARK. Take a group of the form $G = Z_r^n$, where Z_r is a cyclic group of order r, and a set $A \subset G$ of the form

$$A = (a_1 + G') \cup \cdots \cup (a_k + G')$$

with a subgroup G', where the cosets are all disjoint. Here $|A| = m = k|G'|$, and if all the sums $a_i + a_j$ lie in different cosets of G', then

$$|A + A| = \frac{k(k+1)}{2}|G'| = \alpha m, \quad \alpha = \frac{k+1}{2}.$$

The subgroup generated by A can have as many as $r^k|G'|$ elements, hence our function

$$f(r, \alpha) = \alpha^2 r^{\alpha^4}$$

cannot be replaced by anything smaller than

$$r^k = r^{2\alpha - 1}.$$

By recent improvements of the above argument by Green-Ruzsa and then Sanders, the above bound is now almost achieved.

The following conjecture of Katalin Marton would yield a more efficient covering in a slightly different form.

CONJECTURE 2.2. If $|A| = n$, $|A + A| \leq \alpha n$, then there is a subgroup H of G such that $|H| \leq n$ and A is contained in the union of α^c cosets of H, where the constant c may depend on r but not on n or α.

In the most optimistic form c would be $1 + o(1)$..

This is equivalent to the following problem, which I think is interesting in its own right.

CONJECTURE 2.3 (Equivalent conjecture.). Let G be as above, $f : G \to G$ a function such that $f(x + y) - f(x) - f(y)$ assumes at most α distinct values. Then f has a decomposition $f = g + h$, where g is a homomorphism and h assumes $\leq \alpha^c$ values.

The equivalence is meant in a loose sense, the values of c need not be the same. (The proof of this equivalence is unpublished.)

3. Freiman isomorphism and small models

DEFINITION 3.1. Let G_1, G_2 be commutative groups, $A_1 \subset G_1$, $A_2 \subset G_2$. We say that a mapping $\varphi : A_1 \to A_2$ is a *homomorphism of order r in the sense of Freiman*, or an *F_r-homomorphism* for short, if for every $x_1, \dots, x_r, y_1, \dots, y_r \in A_1$ (not necessarily distinct) the equation

$$x_1 + x_2 + \cdots + x_r = y_1 + y_2 + \cdots + y_r \tag{3.1}$$

implies

$$\varphi(x_1) + \varphi(x_2) + \cdots + \varphi(x_r) = \varphi(y_1) + \varphi(y_2) + \cdots + \varphi(y_r). \tag{3.2}$$

We call φ an *F_r-isomorphism*, if it is (1-1) and its inverse is a homomorphism as well, that is, (3.2) holds if and only if (3.1) does. If we say Freiman-homomorphism or isomorphism without specifying r, then the first nontrivial case $r = 2$ is meant.

Any affine linear function is an F_r-isomorphism for every r, and the nondegenerate ones are F_r-isomorphisms.

PREXERCISE. If one of two F-isomorphic sets contains an l-term arithmetic progression, then so does the other.

PREXERCISE. If A and B are F_r-isomorphic with $r = q(k + l)$, then $kA - lA$ and $kB - lB$ are F_q-isomorphic.

PREXERCISE. The F-homomorphic image of a d-dimensional arithmetic progression is also a d-dimensional arithmetic progression with the same "lengths" $l_1, .., l_d$.

A Freiman isomorphism preserves additive properties up to a point. We show that being a generalized arithmetic progression is such a property.

LEMMA 3.2. *Let G, G' be commutative groups. If a set $P' \subset G'$ is the homomorphic image of a generalized arithmetical progression $P(q_1, \dots, q_d; l_1, \dots, l_d; a) \subset G$, then there are elements $q_1', \dots, q_d', a' \in G'$ such that*

$$P' = P(q_1', \dots, q_d'; l_1, \dots, l_d; a') \tag{3.3}$$

and the homomorphism is given by

$$\phi(a + x_1 q_1 + \cdots + x_d q_d) = a' + x_1 q_1' + \cdots + x_d q_d'. \tag{3.4}$$

PROOF. Define a' and q_i' by

$$a' = \phi(a), \quad q_i' = \phi(a + q_i) - \phi(a).$$

We prove (3.4) by induction on $r = x_1 + \cdots + x_d$. For $r \le 1$ it is an immediate consequence of the definition. Assume that $r \ge 2$ and the statement holds for every smaller value. Consider an element

$$x = x_1 q_1 + \cdots + x_d q_d, \quad x_1 + \cdots + x_d = r.$$

Since $r \ge 2$, either there are subscripts $i \ne j$ such that $x_i \ge 1$ and $x_j \ge 1$, or there is a subscript for which $x_i \ge 2$. In the second case write $j = i$. In both cases the sums

$$y = x - x_i, \ z = x - x_j, \ u = x - x_i - x_j$$

are in P, their sums of coefficients are at most $r - 1$ and they satisfy $x + u = y + z$. This implies $\phi(x) + \phi(u) = \phi(y) + \phi(z)$, that is, $\phi(x) = \phi(y) + \phi(z) - \phi(u)$. Substituting (3.4) for y, z and u into this equation we conclude that (3.4) holds for x as well, which completes the inductive step. □

LEMMA 3.3. *Let G, G' be commutative groups, and let $A \subset G$, $A' \subset G'$ be F_r-isomorphic sets. Assume that $r = r'(k+l)$ with nonnegative integers r', k, l. The sets $kA - lA$ and $kA' - lA'$ are $F_{r'}$-isomorphic.*

PROOF. Let ϕ be the isomorphism between A and A'. For an

$$x \in kA - lA,\ x = a_1 + \cdots + a_k - b_1 - \cdots - b_l$$

we define naturally

$$\psi(x) = \phi(a_1) + \cdots + \phi(a_k) - \phi(b_1) - \cdots - \phi(b_l).$$

The facts that this depends only on x and not on the particular representation, and that ψ is an $F_{r'}$-isomorphism, follow immediately from the definition. □

With this concept we can formulate principle (iii) from the Introduction of Chapter 1 exactly.

LEMMA 3.4. *Let A be a finite set in a torsionfree commutative group, and let r be any positive integer. There is a set $A' \subset \mathbb{Z}$ which is F_r-isomorphic to A.*

The proof, as also outlined there, consists of first applying the structure theorem of finitely generated torsionfree groups to reduce the general case to sets lying in $\mathbb{Z}^d$, and then a suitable projection to go to $\mathbb{Z}$.

We define the *Freiman dimension* of a set $A \subset \mathbb{R}^k$ as the largest d for which there is an isomorphic properly d-dimensional set.

EXERCISE 22. For a set $A \subset \mathbb{Z}^d$ the following are equivalent:
a) its Freiman dimension is d,
b) every Freiman homomorphism from A to any $\mathbb{R}^k$ is affine linear.

The first step towards finding the structure of a set will be to find a Freiman-isomorphic image, or "model", which is comfortably sitting in a small group or interval.

THEOREM 3.5. *Let A be a finite set in a torsionfree commutative group, $|A| = m$, $r \geq 2$ an integer and $|rA - rA| = n$.*
(a) For every $q \geq n$ there exists a set $A' \subset A$, $|A'| \geq m/r$ which is F_r-isomorphic to a set T' of residues modulo q.
(b) There is a set $A^* \subset A$, $|A^*| \geq m/r^2$ which is F_r-isomorphic to a set T^* of integers,

$$T^* \subset [0, n/r].$$

PROOF. In view of the previous lemma we may assume that $A \subset \mathbb{Z}$.

The isomorphism in (a) will be given by a function

$$\varphi(a) = [\xi a] \pmod q$$

for a suitably chosen real number $\xi \in [0, q]$, and the set A' will be one of the r sets

$$A_j = \left\{ a \in A : \frac{j-1}{r} \leq \{\xi a\} < \frac{j}{r} \right\},\ j = 1, \ldots, r.$$

We claim that for a suitable choice of ξ the restriction of φ is an isomorphism on each set A_j; clearly at least one of them will have $\geq m/r$ elements.

This isomorphism means that for arbitrary $a_1, \ldots, a_r, b_1, \ldots, b_r \in A_j$ the congruence

$$[\xi a_1] + \cdots + [\xi a_r] \equiv [\xi b_1] + \cdots + [\xi b_r] \pmod q \tag{3.5}$$

should be equivalent to the equality

$$a_1 + \cdots + a_r = b_1 + \cdots + b_r.$$

First we show that this equality implies

$$[\xi a_1] + \cdots + [\xi a_r] = [\xi b_1] + \cdots + [\xi b_r],$$

and a fortiori the congruence (3.5) for every ξ. Indeed,

$$\sum\big([\xi a_i] - [\xi b_i]\big) = \xi \sum (a_i - b_i) - \sum\big(\{\xi a_i\} - \{\xi b_i\}\big). \tag{3.6}$$

If all the fractional parts are in an interval $[u, u + 1/r)$, then the absolute value of the last sum is < 1. The left side, as an integer with absolute value < 1, must be 0.

Assume now congruence (3.5). The left side of (3.6) is a multiple of q, and the right side is of the form $\xi t + \delta$, where $t \in rA - rA$ and $|\delta| < 1$. We want to infer $t = 0$, that is, we try to exclude all possible equalities of the type $kq = \xi t + \delta$, or

$$\xi = \frac{kq - \delta}{t}$$

wit For a given value of t this is a collection of $t + 1$ intervals of total length 2. If the union of these systems of intervals does not cover $[0, q]$, we can find a ξ which is not contained in any of them. The number of values of t that we have to take into account is $(n - 1)/2$, since t and $-t$ induce the same collection of excluded intervals. Hence a sufficient condition is $2(n - 1)/2 < q$, or $q \geq n$.

To prove part (b), we combine this map φ with $\psi : \mathbb{Z}_q \to \mathbb{Z}$, where ψ is the smallest nonnegative representation of a residue class. We split the integers of the interval $[0, q - 1]$ into r almost equal subintervals of type $[(i - 1)q/r, iq/r)$, $i = 1, \ldots, r$. The r-fold sums from a fixed interval lie in an interval of length $< q$, thus they are incongruent modulo q unless they are equal. This division splits A' into r parts, and any can serve as A^*. In this way we can achieve

$$|A^*| \geq |A'|/r \geq m/r^2.$$

The isomorphic image of A^* lies in an interval of type $[(i-1)q/r, iq/r)$, and a shift takes it into $[0, q/r]$. For q we take the smallest guaranteed value $q = n$. □

Exercise 23. Let p be a prime, $A \subset \mathbb{Z}_p$, $|A| = n$, k a positive integer. If $p > k^n$, then there is a $t \in \mathbb{Z}_p$, $t \neq 0$ such that $||at/p|| \leq 1/k$ for all $a \in A$.

Exercise 24. Let $A \subset \mathbb{N}$, $|A| = n$. Prove that there is a Freiman isomorphic set contained in $[0, 4^n]$.

Exercise 25. Show that the bound in the previous exercise cannot be improved below 2^{n-2}.

4. Elements of Fourier analysis on groups

In this section we collect some basic facts about the Fourier transform which will be used in the next section. Detailed proofs are not given; instead the main statements are split into several exercises, which even the uniniatiated reader may try to solve. It is not *necessary* to solve these exercises to understand the next section; the prerequisites are here in the form of definitions and statements, but it certainly helps.

A group will mean a commutative group; a *character* is a homomorphism $\gamma : G \to \mathbb{C}_1$, where $\mathbb{C}_1 = \{z : |z| = 1\}$ (with multiplication). So if the operation in G is

denoted additively, then $\gamma(x+y) = \gamma(x)\gamma(y)$. The characters of G form a group (under pointwise multiplication). We write mostly Γ to denote this group. Its unity is $\gamma_0 \equiv 1$, the *principal character*. We write $\bar{\gamma}(g) = \overline{\gamma(g)}$; it is the inverse, and at the same time the pointwise complex conjugate of γ.

Characters of a cyclic group $\mathbb{Z}_q$ are simple. Indeed, if γ is a character and $\gamma(1) = \omega$, then $\gamma(n) = \omega^n$. Since $\gamma(q) = \gamma(0) = 1$, we see that ω must be a q-th root of unity, say $\omega = e^{2\pi ik/q}$ with some k, consequently

$$\gamma(n) = e^{2\pi ikn/q}.$$

If we restrict our attention to cyclic groups, which is the most important object for the sequel, then we could just use the above functions and not even mention the word "character". I think, however, that this is the natural way of presentation (some reasons are given later).

The above formula shows that $\mathbb{Z}_q$ has exactly q characters, moreover they also form a cyclic group of order q. This is not so obvious for other groups.

We write $G_1 < G$ to denote that G_1 is a subgroup of G.

Exercise 26. Let $G_1 < G$, $g \in G \backslash G_1$, γ a character of G_1. γ can be extended to a character of the group G_2 generated by $G_1 \cup \{g\}$.

Exercise 27. This γ above can be extended to a character of G. Consequently, for any $g \in G$, $g \neq e$ (e =unity) there is a character γ with $\gamma(g) \neq 1$ (in other words, the charaters separate G).

Exercise 28. The only important property of $\mathbb{C}_1$ in the above exercises is that it is *divisible*. A group G is divisible if for every $g \in G$ and positive integer k there is an $h \in G$ such that $h^k = g$. (We use multiplicative notation here for compatibility with $\mathbb{C}_1$.) Show that the previous exercises hold with the set of homomorphisms to any fixed divisible group in the place of $\mathbb{C}_1$.

Exercise 29. Extend the previous three exercises to infinite groups.

Exercise 30. $\sum_{g\in G} \gamma(g) = 0$ unless $\gamma = \gamma_0 \equiv 1$. Hint: compare it to $\sum_{g\in G} \gamma(ag)$.

Exercise 31. $\sum_{\gamma\in\Gamma} \gamma(g) = 0$ unless $g = 0$.

Exercise 32. $|\Gamma| = |G|$. Hint: consider $\sum_{g\in G}\sum_{\gamma\in\Gamma} \gamma(g)$.

Exercise 33. For a $g \in G$, we define a character g^* of Γ by $g^*(\gamma) = \gamma(g)$. The mapping $g \to g^*$ embeds G into $\hat{\Gamma}$, the group of characters of Γ. Show that this is an isomorphism.

Exercise 34. The previous exercise fails for infinite groups. In fact, it is wrong for each infinite group.

Exercise 35. If $G = G_1 \times G_2$, then Γ is isomorphic to $\Gamma_1 \times \Gamma_2$.

Exercise 36. For finite groups, Γ is isomorphic to G.

Exercise 37. The previous exercise also provides a direct access exercises 32 and 33. It is of limited value, since this isomorphy is not natural: we cannot find a way to define a 1–1 correspondence between G and Γ. Try to formulate this observation exactly and then prove it.

EXERCISE 38. For functions $\alpha, \beta : G \to \mathbb{C}$ we define a direct product by

$$(\alpha, \beta) = |G|^{-1} \sum_{g \in G} \alpha(g)\beta(\tilde{g}).$$

This turns Γ into an orthonormal system: for $\gamma, \gamma' \in \Gamma$ we have $(\gamma, \gamma') = 0$ if $\gamma \neq \gamma'$, $=1$ if $\gamma = \gamma'$.

EXERCISE 39. Every funtion α on G has a development into a character series $\alpha = \sum_{\gamma \in \Gamma} c_\gamma \gamma$. Express the coefficients c_γ. Find the development of the indicator function of an element.

DEFINITION 4.1. Let $\varphi : G \to \mathbb{C}$ be a function on the group G. Its *Fourier transform* is the function $f : \Gamma \to \mathbb{C}$ defined by

$$f(\gamma) = \sum_{g \in G} \varphi(g)\gamma(g).$$

The Fourier transform is often denoted by $f = \hat{\varphi}$

For a cyclic group $G = \mathbb{Z}_q$ the characters are the functions

$$\gamma_k(n) = e^{2\pi ik/n}, \ k = 0, 1, \ldots, q-1.$$

Consequently the Fourier transform of a function φ is given by

$$f(\gamma_k) = \sum_n e^{2\pi ikn/q}\varphi(n).$$

If we identify this character γ_k with its subscript $k \in \mathbb{Z}_q$, we can also say that the Fourier transform is

$$f(k) = \sum_n e^{2\pi ikn/q}\varphi(n),$$

which is frequently done when no other group is used. In this booklet we will distinguish G and Γ, elements and characters, for methodological reasons.

Given the Fourier transform of a function, we can reconstruct the function from it as follows.

STATEMENT 4.2 (Fourier inversion formula.). *Let φ be a function on G and $f = \hat{\varphi}$ its Fourier transform. We have*

$$\varphi(x) = \frac{1}{|G|} \sum_{\gamma \in \Gamma} f(\gamma)\overline{\gamma}(x).$$

EXERCISE 40. Prove the inversion formula. (Is this a new exercise or an old one?)

EXERCISE 41. How does the inversion formula look for the group $\mathbb{Z}_q$?

Another important fact is the analog of the Parseval (or Plancherel) identity.

STATEMENT 4.3 (Parseval formula.). *Let φ be a function on G and $f = \hat{\varphi}$ its Fourier transform. We have*

$$\sum_{\gamma \in \Gamma} |f(\gamma)|^2 = |G| \sum_{x \in G} |\varphi(x)|^2 .$$

EXERCISE 42. Prove the Parseval formula.

EXERCISE 43. Let φ_1, φ_2 be functions on G, with Fourier transforms f_1, f_2. What is the connection between the direct products (φ_1, φ_2) and (f_1, f_2)?

The case of 0-1 valued functions is of special imporance for us. Let $A \subset G$ be any set, and consider its *indicator function*

$$\varphi(x) = \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{if } x \notin A. \end{cases}$$

Its Fourier transform is

$$f(\gamma) = \sum_{a \in A} \gamma(a). \tag{4.1}$$

With an abuse of terminology we shall call this the *Fourier transform of the set* A and denote it by $\hat{A}(\gamma)$.

EXERCISE 44. What does the Parseval formula tell for the Fourier transform of a set?

EXERCISE 45. What does the inversion formula tell for the Fourier transform of a set?

EXERCISE 46. If the Fourier transform of a set A is f, what is the transform of the set $-A$?

Let now A_1, A_2 be sets in G with Fourier transforms f_1, f_2. By using the definition (4.1) and multiplying we obtain

$$f_1(\gamma) f_2(\gamma) = \sum_{x \in G} r(x),$$

where

$$r(x) = \left| \{(a_1, a_2) : a_i \in A_i, a_1 + a_2 = x\} \right|,$$

the number of representations of x as a sum with summands from our sets. The inversion formula now gives

$$r(x) = \frac{1}{|G|} \sum_{\gamma \in \Gamma} f_1(\gamma) f_2(\gamma) \overline{\gamma}(x),$$

and in principle this gives a complete description of the sumset. This is the basis of the usage of analytic methods in additive number theory (under various names, like generating functions, circle method, Hardy-Littlewood method, depending on particular appearances).

EXERCISE 47. If the Fourier transform of a set A is f, whose transform is $|f|^2$?

For the next exercises let A be a set of integers, $|A| = n$ and

$$\hat{A}(t) = \sum_{a \in A} e^{2\pi i a t}, \ t \in \mathbb{R}.$$

EXERCISE 48. What is the connection between this function $\hat{A}$ of a real variable for $A \subset \mathbb{Z}$ and the function $\hat{A}(\gamma)$ for $A \subset \mathbb{Z}_q$?

EXERCISE 49. What is the arithmetical meaning of the integral $\int_0^1 |\hat{A}(t)|^4 dt$? What is its minimal value?

EXERCISE 50. What is the maximal value of the integral in the previous exercise, and for which sets does it occur?

EXERCISE 51. How can one express the number of three-term arithmetical progressions in A (that is, the number of pairs a, d such that $a, a+d, a+2d \in A$) by the function $\hat{A}$?

EXERCISE 52. And what happens if we count only those where $d > 0$?

5. Bohr sets in sumsets

DEFINITION 5.1. If G is a commutative group, $\gamma_1, \dots, \gamma_k$ are characters of G and $\varepsilon_j > 0$, we write

$$B(\gamma_1, \dots, \gamma_k; \varepsilon_1, \dots, \varepsilon_k) = \{g \in G : |\arg \gamma_j(g)| \leq 2\pi\varepsilon_j \text{ for } j = 1, \dots, k\}$$

and call these sets *Bohr sets*. In particular, if $\varepsilon_1 = \dots = \varepsilon_k = \varepsilon$, we shall speak of a *Bohr* (k, ε)*-set*. (We take the branch of arg that lies in $[-\pi, \pi)$.)

In locally compact groups these sets form a base for the Bohr topology; we shall work with finite groups, but we preserve the name that suggests certain ideas.

We shall work mainly with the simplest possible cyclic groups $\mathbb{Z}_q$. Here a typical character is of the form

$$\gamma(x) = e^{2\pi i u x/q}, \ u \in \mathbb{Z}_q,$$

so $\arg \gamma(x) = 2\pi \|ux/q\|$, where $\|t\| = \min(\{t\}, 1 - \{t\})$ denotes the *absolute fractional part* of t, its distance from the nearest integer. In these formulas we were tacitly cheating a bit; for $u, x \in \mathbb{Z}_q$ we replaced them by any integer in the corresponding residue class, and though ux/q can have many different values, it is unique modulo one, so the fractional part and the exponential are uniquely determined.

Hence a Bohr set in $\mathbb{Z}_q$ can be written as

$$B(u_1, \dots, u_k; \varepsilon_1, \dots, \varepsilon_k) = \{x \in \mathbb{Z}_q : \|u_j x/q\| \leq \varepsilon_j \text{ for } j = 1, \dots, k\}.$$

We shall see in the next section that Bohr sets are rather similar to multidimensional arithmetic progressions.

LEMMA 5.2. *Let G be a finite commutative group, $|G| = q$. Let A be a nonempty subset of G and write $|A| = m = \beta q$. The set $D = 2A - 2A$ (the second difference set of A) contains a Bohr (k, ε)-set with some integer $k < \beta^{-2}$ and $\varepsilon = 1/4$.*

This is essentially a result of Bogolyubov [**3**] which he used to study the Bohr topology on the integers.

PROOF. Let Γ denote the group of characters. For $\gamma \in \Gamma$ put

$$f(\gamma) = \sum_{a \in A} \gamma(a).$$

We have

$$\sum_{\gamma \in \Gamma} |f(\gamma)|^2 = mq = \beta q^2$$

(Parseval formula) and $f(\gamma_0) = m$ for the principal character $\gamma_0 (\equiv 1)$.

Recall that $\overline{f(\gamma)}$ is the series corresponding to the set $-A$. Multiplying two copies of f and two copies of $\overline{f}$ we find that

$$|f(\gamma)|^4 = \sum r(x)\gamma(x),$$

where $r(x)$ counts the quadruples $a_1, a_2, a_3, a_4 \in A$ such that $a_1 + a_2 - a_3 - a_4 = x$. A Fourier inversion now gives

$$r(x) = \frac{1}{q}\sum_{\gamma\in\Gamma} |f(\gamma)|^4\gamma(x).$$

Therefore we have $x \in D$ for those elements x for which

$$\sum_{\gamma\in\Gamma} |f(\gamma)|^4\gamma(x) \neq 0. \tag{5.1}$$

To estimate (5.1), we split the characters $\gamma \neq \gamma_0$ into two groups. We put those for which $|f(\gamma)| \geq \sqrt{\beta}q$ into Γ_1 and the rest into Γ_2. We claim that $x \in D$ whenever $\operatorname{Re}\gamma(x) \geq 0$ is satisfied for all $\gamma \in \Gamma_1$. Indeed, we have

$$\left|\sum_{\gamma\in\Gamma_2} |f(\gamma)|^4\gamma(x)\right| < \beta q^2 \sum_{\gamma\in\Gamma_2} |f(\gamma)|^2 < \beta^2 m^2 q^2 = m^4,$$

consequently

$$\operatorname{Re}\sum_{\gamma\in\Gamma} |f(\gamma)|^4\gamma(x) \geq m^4 + \operatorname{Re}\sum_{\gamma\in\Gamma_2} |f(\gamma)|^4\gamma(x) \geq m^4 - \left|\sum_{\gamma\in\Gamma_2} |f(\gamma)|^4\gamma(x)\right| > 0.$$

The condition $\operatorname{Re}\gamma(x) \geq 0$ is equivalent to $|\arg\gamma(g)| \leq \pi/2$, thus we have a Bohr $(k, 1/4)$ set with $k = |\Gamma_1|$. We estimate k. We have

$$k\beta m^2 \leq \sum_{\gamma\in\Gamma_1} |f(\gamma)|^2 < \sum_{\gamma\in\Gamma} |f(\gamma)|^2 = \beta q^2,$$

hence $k \leq (q/m)^2 = \beta^{-2}$ as claimed. □

This theorem used four copies of the set A. A similar result as Theorem 5.2 holds for 3 sets, even for different ones.

THEOREM 5.3. *If A_1, A_2, A_3 are subsets of G, a commutative group with $|G| = q$ and $|A_i| \geq \beta_i q$ then, for some t, $A_1 + A_2 + A_3 \supset t + B(\gamma_1, \ldots, \gamma_k, \eta)$, where k and η depend only on the densities β_i.*

The corresponding result for two copies does not hold, not even for the difference set $A - A$. The reason for this is the following. A Bohr set always contains a long arithmetic progression. This will be proved in a stronger form in the next section.

PREXERCISE. A Bohr (k, ε) set in $\mathbb{Z}_q$ contains an arithmetical progression of length n^δ, where $\delta > 0$ depends on k and ε.

However, the set $A - A$ may not contain an arithmetic progression of length q^δ, with $\delta = \delta(\beta)$, assuming $|A| \geq \beta n$. The maximal length of the arithmetic progression may be $< e^{\log q^{2/3+\epsilon}}$. On the other hand it is known that it is $\gg e^{\log q^{1/2-\epsilon}}$ (Green and Bourgain).

6. Some facts from the geometry of numbers

We consider sets situated in an Euclidean space $\mathbb{R}^d$.

DEFINITION 6.1. A set $L \subset \mathbb{R}^d$ is a *lattice*, if it is a discrete subgroup and it is not contained in any smaller dimensional subspace.

Any such lattice is necessarily isomorphic to $\mathbb{Z}^d$, that is, there are linearly independent vectors $e_1, \dots, e_d \in \mathbb{R}^d$ such that

$$L = \{x_1 e_1 + \dots + x_d e_d : x_i \in \mathbb{Z}\}.$$

DEFINITION 6.2. A set $F \subset \mathbb{R}^d$ is a *fundamental domain* of this lattice if the sets $F + x$, $x \in L$ cover R^d without overlap (one representant from each coset of L). (Sometimes overlaps of boundaries is permitted.)

An example is

$$F = \{x_1 e_1 + \dots + x_d e_d : 0 \le x_i < 1\}.$$

EXERCISE 53. Prove that measurable fundamental domains all have the same volume.

On the example of the domain above one can see that this is the absolute value of the determinant formed by the vectors e_i, which is hence independent of the choice of the basis (e_i).

DEFINITION 6.3. The common value of volumes of fundamental domains and absolute value of determinants of matrices formed by integral bases is called the *determinant of the lattice.*

EXERCISE 54. If $L \subset \mathbb{Z}^d$ is a lattice, its determinant is the same as its index in $\mathbb{Z}^d$ as a subgroup.

DEFINITION 6.4. Let Q be a closed neighbourhood of 0, and let L be a lattice in $\mathbb{R}^d$. The *successive minima* of Q with respect to the lattice are the smallest positive numbers $0 < \lambda_1 \le \dots \le \lambda_d$ such that there are linearly independent vectors $a_1, \dots, a_d \in L$, $a_i \in \lambda_i Q$.

Imagine this as follows. Take a small homothetic image εQ and blow it up slowly. First the only lattice point inside is the origin, then at λ_1 another appears. As we increase λ, it may happen that the next lattice points are multiples of a_1, like $2a_1$ at $2\lambda_1$, but at some point λ_2 we get another, which is not a multiple of the first and so on.

EXERCISE 55. Show that the first appearing vectors a_i may not form a basis of L.

We will need the following important theorem of Minkowski.

LEMMA 6.5 (Minkowski's inequality for successive minima.). *Let Q be a closed neighbourhood of 0, and let L be a lattice in $\mathbb{R}^d$. Let $0 < \lambda_1 \le \dots \le \lambda_d$ be the successive minima of Q with respect to L. We have*

$$\lambda_1 \dots \lambda_d \le 2^d \frac{\det L}{\operatorname{vol} Q}. \tag{6.1}$$

7. A generalized arithmetical progression in a Bohr set

We show that Bohr sets contain large generalized arithmetical progressions. We will do this for cyclic groups only; for general groups see Green and Ruzsa [...]

THEOREM 7.1. *Let q be a positive integer, $u_1, \dots, u_d$ residues modulo q such that $(u_1, u_2, \dots, u_d, q) = 1$, $\varepsilon_1, \dots, \varepsilon_d$ real numbers satisfying $0 < \varepsilon_j < 1/2$. Write*

$$\delta = \frac{\varepsilon_1 \dots \varepsilon_d}{d^d}. \tag{7.1}$$

There are residues $v_1, \dots, v_d$ and nonnegative integers $l_1, \dots, l_d$ such that the set

$$P = \{v_1 x_1 + \dots v_d x_d : |x_i| \le l_i\} \tag{7.2}$$

satisfies

$$P \subset B(u_1, \dots, u_d; \varepsilon_1, \dots, \varepsilon_d), \tag{7.3}$$

the sums in (7.2) are all distinct and

$$|P| = \|P\| = \prod (2l_j + 1) \ge \prod (l_j + 1) > \delta q. \tag{7.4}$$

PROOF. Let L be the d dimensional lattice of integer vectors $(x_1, \dots, x_d)$ satisfying

$$x_1 \equiv x u_1, \dots, x_d \equiv x u_d \pmod q$$

with some integer x. This lattice is the union of q translations of the lattice $(q\mathbb{Z})^d$ (here we need the coprimality condition, otherwise there may be coincidences), hence its determinant is q^{d-1}.

Let Q be the rectangle determined by $|x_j| \le \varepsilon_j$, $j = 1, \dots, d$ and let $\lambda_1, \dots, \lambda_d$ denote the successive minima of Q with respect to the lattice L. These are the smallest positive numbers such that there are linearly independent vectors $a_1, \dots, a_d \in L$, $a_i \in \lambda_i Q$. By Minkowski's inequality (6.1) we have

$$\lambda_1 \dots \lambda_d \le 2^d \frac{\det L}{\operatorname{vol} Q} = \frac{q^{d-1}}{\varepsilon_1 \dots \varepsilon_d}. \tag{7.5}$$

Write

$$a_i = (a_{i1}, \dots, a_{id}).$$

The condition $a_i \in \lambda_i Q$ means that $|a_{ij}| \le \lambda_i \varepsilon_j$. Since $a_i \in L$, there are residues v_i such that $a_{ij} \equiv v_i u_j \pmod q$. These are our v_j's and we put

$$l_i = \left[\frac{q}{d\lambda_i} \right].$$

First we show that $P \subset B$. Consider an $x \in P$, $x = x_1 v_1 + \dots + x_d v_d$. We have

$$x u_j = \sum x_i v_i u_j \equiv \sum x_i a_{ij} \pmod q,$$

consequently

$$\begin{aligned}\left\|\frac{xu_j}{q}\right\| &= \left\|\sum \frac{x_i a_{ij}}{q}\right\| \\ &\le \sum \left|\frac{x_i a_{ij}}{q}\right| \\ &\le \sum \frac{l_i \lambda_i \varepsilon_j}{q} \le \sum \frac{\varepsilon_j}{d} = \varepsilon_j.\end{aligned} \tag{7.6}$$

Next we show that these elements are all distinct. If $x_1, \dots, x_d$ and $y_1, \dots, y_d$ give the same sum, then with $z_j = x_j - y_j$ we have

$$\sum z_i v_i \equiv 0 \pmod q, \quad |z_i| \le 2l_i.$$

Multiplying this congruence by u_j we infer that

$$\sum z_i a_{ij} \equiv 0 \pmod q$$

for all j. Moreover a calculation like above yields

$$\left|\sum z_i a_{ij}\right| \le \sum l_i \lambda_i \varepsilon_j \le 2\varepsilon_j q < q.$$

Consequently $\sum z_i a_{ij} = 0$ for every j, which means that $\sum z_i a_i = 0$; by view of the linear independence of the vectors a_i, $z_i = 0$ for all i, q. e. d.

Finally we prove (7.4). We have

$$l_i + 1 > \frac{q}{d\lambda_i},$$

hence

$$\prod (l_i + 1) > \frac{q^d}{d^d \lambda_1 \dots \lambda_d} \ge \frac{q}{d^d} \varepsilon_1 \dots \varepsilon_d = \delta q$$

by (7.5). □

It is easy to see that the result need not hold if $(u_1, \dots, u_d, q) > 1$; consider, for instance, the case $q = r^2$, $d = 1$, $u_1 = r$. It can be shown that a $d + 1$ dimensional arithmetical progression can always be found in B.

LEMMA 7.2. *Let q be a prime, and let A be a nonempty set of residues modulo q with $|A| = \beta q$. There are residues $v_1, \dots, v_d$ and nonnegative integers $l_1, \dots, l_d$ such that the set*

$$P = \{v_1 x_1 + \dots v_d x_d : |x_i| \le l_i\} \tag{7.7}$$

satisfies $P \subset D = 2A - 2A$, the sums in (7.7) are all distinct and

$$\|P\| = \prod (2l_j + 1) \ge \prod (l_j + 1) > \delta q. \tag{7.8}$$

where $d \le \beta^{-2}$ and

$$\delta = (4d)^{-d} \le (\beta^2/4)^{1/\beta^2}. \tag{7.9}$$

PROOF. This follows from a combination of Lemma 5.2 and Theorem 7.1. The assumption that q is a prime guarantees the coprimality assumption required in Theorem 3.1. □

8. Freiman's theorem

We prove Freiman's Theorem 1.3 in the following form.

THEOREM 8.1. *Let A, B be finite sets in a torsionfree commutative group satisfying $|A| = |B| = m$, $|A + B| \le \alpha m$. There are numbers d, s depending on α only such that A is contained in a generalized arithmetical progression of dimension at most d and size at most sm.*

Since a bound on $A + B$ immediately gives a bound on $2A$, the generalization to different sets is not important as long as we do not give bounds for d and s. Given a set in a torsionfree group we can find Freiman-isomorphic sets in $\mathbb{Z}$, however, it is not completely obvious (though not very difficult) to deduce the existence of a covering progression from that of an isomorphic image. The form above is just the natural one in our treatment.

PROOF. We apply Theorem 3.5 for $r = 8$ and a prime number $q > |rA - rA|$. By Chebyshev's theorem we can find such a prime with

$$q < 2|rA - rA| \le 2\alpha^{16} m;$$

the second inequality follows from Theorem 1.1 of Chapter 1. We obtain a set $A' \subset A$, which is F_8-isomorphic to a set T of residues modulo q, $|A'| \ge m/r = m/8$.

Applying Lemma 7.2 we find a d' dimensional proper arithmetical progression $P \subset 2T - 2T$ of size $\ge \delta m$, where $d' = d'(\alpha)$ and $\delta = \delta(\alpha) > 0$ depend on α only.

By Lemma 3.3 the F_8-isomorphism between T and A' induces an F_2-isomorphism between $2T - 2T$ and $2A' - 2A'$. The image P' of P is a proper d'-dimensional arithmetical progression by Lemma 3.2 and we have $P' \subset 2A' - 2A' \subset 2A - 2A$.

Select a maximal collection of elements $a_1, \dots, a_t \in A$ such that the sets $P' + a_i$ are pairwise disjoint. We estimate t. Since these sets are all subsets of $A + P' \subset 3A - 2A$, we have

$$t \le \frac{|3A - 2A|}{\|P'\|} \le \frac{\alpha^5 m}{\delta m} = \alpha^5/\delta(\alpha).$$

For every $a \in A$ there is an a_i such that

$$(a + P') \cap (a_i + P') \ne \emptyset.$$

Thus there are $p, p' \in P'$ such that $a + p = a_i + p'$, that is, $a = a_i + p' - p$. This means that

$$A \subset \{a_1, \dots, a_t\} + P' - P'. \tag{8.1}$$

Since P' is a d-dimensional arithmetical progression, so is $P' - P'$, and obviously

$$\|P' - P'\| \le 2^d \|P'\| \le 2^d |2A - 2A| \le 2^d \alpha^4 m.$$

The set $\{a_1, \dots, a_t\}$ can be covered by the t-dimensional arithmetical progression

$$P(a_1, \dots, a_t; 1, \dots, 1; 0).$$

Hence the right side of (8.1) can be covered by an arithmetical progression of dimension $d = d' + t$ and size sm, $s = 2^d \alpha^4$. Since both t and d were bounded in terms of α, the proof is completed. □

9. Arithmetic progressions in sets with small sumset

We show that sets with small sumset contain long arithmetic progressions. This will be a (necessarily) conditional result, depending on our knowledge of arithmetic progressions in dense sets. The first such result is again due to Freiman [**10**, Theorem 2.30]. He considered three-term progressions only, since Szemerédi's theorem on long progressions was not yet available.

Let $r_k(n)$ denote the maximal number of integers that can be selected from the interval $[1, n]$ without including a k term arithmetical progression and write

$$\omega_k(n) = n/r_k(n).$$

Szemerédi's celebrated theorem [**59**] tells us that $\omega_k(n) \to \infty$ for every fixed k. The best known estimates are due to Gowers [**13, 14**] for general k, and to Bourgain [**4**] for $k = 3$.

THEOREM 9.1. *Assume that $|A| = n$ and A does not contain any k-term arithmetical progression. We have*

$$|A + A - A - A| \geq \frac{1}{4}\omega_k(n)n, \tag{9.1}$$

$$|A + B| \geq \frac{1}{\sqrt{2}}\omega_k(n)^{1/4}n^{1/4}|B|^{3/4} \tag{9.2}$$

for every set B,

$$|A + B| \geq \frac{1}{\sqrt{2}}\omega_k(n)^{1/4}n \tag{9.3}$$

for every set B such that $|B| = n$,

$$|A + A| \geq \frac{1}{\sqrt{2}}\omega_k(n)^{1/4}n, \tag{9.4}$$

$$|A - A| \geq \frac{1}{\sqrt{2}}\omega_k(n)^{1/4}n. \tag{9.5}$$

By Bourgain's result we have $\omega_3(n) \gg (\log n)^{1/2-\varepsilon}$. Applying this estimate we obtain the following version of Freiman's theorem.

COROLLARY 9.2. *Assume that $|A| = n$ and A does not contain any 3-term arithmetical progression. For every constant $c < 1/8$ and $n > n_0(c)$ we have*

$$|A + B| \geq \frac{1}{2}n(\log n)^c \tag{9.6}$$

for every set B such that $|B| = n$, in particular

$$|A + A| \geq \frac{1}{2}n(\log n)^c, \tag{9.7}$$

$$|A - A| \geq \frac{1}{2}n(\log n)^c. \tag{9.8}$$

PROBLEM 9.3. Can the exponent $1/4$ in (9.4)- (9.5) be improved to 1 or at least to $1 - \varepsilon$?

PROOF. Write $|A| = n$ and $|2A-2A| = \beta n$. We apply the case $r = 2$ of Theorem 3.5, part (b). We get a set $A^* \subset A$, $|A^*| \geq n/4$ which is isomorphic to a set $T \subset [0, \beta n/2]$. By Lemma 3.2 T contains no k-term arithmetical progression.

Since in an interval of length n there can be at most $r_k(n)$ integers without k-term arithmetical progression and the interval $[0, \beta n/2]$ can be covered by $[1 + \beta/2]$ such intervals, we have

$$n/4 \leq |T| \leq [1 + \beta/2]r_k(n) \leq \beta r_k(n),$$

therefore

$$\beta \geq \frac{1}{4}\frac{n}{r_k(n)},$$

which is equivalent to (9.1).

To obtain (9.2) we apply Theorem 1.1, Ch. 1 and (9.1):

$$|A + B| \geq |B|^{3/4}|2A - 2A|^{1/4} \geq \frac{1}{\sqrt{2}}|B|^{3/4}\omega_k(n)^{1/4}n^{1/4}.$$

(9.3) is the case $|B| = n$ of (9.2), while (9.4)- (9.5) are the cases $B = A$ and $B = -A$ of (9.3).

□

CHAPTER 3

Location and sumsets

1. Introduction

This chapter is about questions of the following kind. Assume we have finite sets A, B in a group G. What can we say about $A+B$ if we know the structure of G, or we have some information about how these sets are situated within G? The "what" will be in most cases a lower estimate for the cardinality.

A familiar example is the classical Cauchy-Davenport inequality.

THEOREM 1.1. *Let p be a prime, $A, B \subset \mathbb{Z}_p$ nonempty sets. We have*

$$|A+B| \geq \min(|A|+|B|-1, p).$$

PREXERCISE. Prove the Cauchy-Davenport inequality by comparing $|A+B|$ and $|A'+B'|$ for suitably chosen sets of the form

$$A' = A \cup (B+t), \quad B' = B \cap (A-t).$$

For another example consider Freiman's Theorem 1.1 from Chapter 2: If $A \subset \mathbb{Z}$, $|A| = m$, $|A+A| \leq 3m-4$, then A is contained in an arithmetic progression of length $\leq |A+A| - m + 1 \leq 2m-3$.

DEFINITION 1.2. The *reduced diameter* $\operatorname{diam} A$ of a set $A \subset \mathbb{Z}$ is the smallest u such that A is contained in an arithmetic progression $\{b, b+q, \dots, b+uq\}$. (Later we shall generalize and rename this concept.)

Now we can formulate Freiman's theorem equivalently as follows.

THEOREM 1.3. *For any set $A \subset \mathbb{Z}$ with $|A| = m$ and $\operatorname{diam} A = u$ we have*

$$|2A| \geq \min(m+u, 3m-3).$$

This illustrates that the distinction between "direct" and "inverse" or "structural" results is often only a case of style. This chapter will contain results that are more naturally expressed in the "direct" form.

First we consider finite groups, then lattices $\mathbb{Z}^d$, then more general structures.

2. The Cauchy-Davenport inequality

Here we give a proof of Theorem 1.1. Several proofs are known, we present a well-known one as outlined in the prexercise above, mainly for the sake of presenting a method in the simplest form which will be used several times later.

This is based on two transformations:

(1) Translation. If we replace A, B by sets $A+x$, $B+y$, the cardinalities of $A, B, A+B$ remain unchanged.

(2) Transfusion (elements go from A to B). We replace A, B by $A' = A \cap B$ and $B' = A \cup B$. This operation does change the cardinalities but preserves their sum:

$$|A'| + |B'| = |A \cap B| + |A \cup B| = |A| + |B|\,. \tag{2.1}$$

It does not increase the sumset: we have

$$A' + B' \subset A + B. \tag{2.2}$$

This operation yields a new pair of sets if $A \not\subset B$ and $A \cap B \neq \emptyset$.

Proof of Theorem 1.1. Write $|A| = m$, $|B| = n$.

We use induction on m. The case $m = 1$ is obvious. Assume now we know the statement for every pair of sets where $1 \leq |A| \leq m - 1$.

Given a pair of sets A, B with $|A| = m$, we try to make a transfusion. If we get a new pair A', B' with $1 \leq |A'| \leq m-1$, then (2.1) and (2.2) complete the inductive step. If this does not work, we have either $A \subset B$ or $A \cap B = \emptyset$.

Now combine this transfusion with a translation. If it never works, we know that for every x we have either $A + x \subset B$ or $(A + x) \cap B = \emptyset$.

Take now an $y \in A - A$, $y \neq 0$; such an y exists if A has at least two elements. Start with an x such that $(A + x) \cap B \neq \emptyset$ (any $x \in B - A$). Then $A + x \subset B$ by the above dichotomy, and then $(A + x + y) \cap B \neq \emptyset$ again: if $y = a' - a$, then

$$a + x + y = a' + x \in B.$$

By repeating this argument we see that all $x, x+y, x+2y, \ldots$ are in this category. This list contains all elements of $\mathbb{Z}_p$, that is, always $A + x \subset B$. Hence $B = \mathbb{Z}_p$ and the claim holds again evidently. □

A set is *sumfree*, if it has no three elements such that $x + y = z$ (so we exclude $2x = z$ too; if we do not, the following exercises change only minimally).

Exercise 56. What is the size of the largest sumfree subset of $[1, n]$?

Exercise 57. What is the size of the largest sumfree subset of $\mathbb{Z}_p$, p prime?

Exercise 58. Every $A \subset \mathbb{N}$, $|A| = n$ has a sumfree subset of cardinality $\geq n/3$.

Exercise 59. Same problem with $n/3 + 1$ for n sufficiently large. (Bourgain's theorem, extremely difficult.)

Exercise 60. The set of positive integers has no partition into finitely many sumfree parts.

3. Kneser's theorem

We show how to extend the Cauchy-Davenport theorem to composite moduli and general commutative groups. A verbatim extension fails, since $A + A = A$ if A is a subgroup.

Definition 3.1. Let S be a nonempty set in a commutative group G. The *stabilizer* or *group of periods* of S is the set

$$\operatorname{stab} S = \{x \in G : x + S = S\}.$$

(This is clearly a subgroup of G.)

THEOREM 3.2. *Let A, B be finite sets in a commutative group G, $S = A + B$ and $H = \operatorname{stab} S$. We have*

$$|A + B| \geq |A + H| + |B + H| - |H|. \tag{3.1}$$

If (3.1) *holds with strict inequality, then*

$$|A + B| \geq |A + H| + |B + H| \geq |A| + |B|. \tag{3.2}$$

This clearly implies the Cauchy-Davenport theorem, as in $\mathbb{Z}_p$ the only possibilities are $H = \{0\}$ or $H = \mathbb{Z}_p$.

LEMMA 3.3. *Let S be a finite set in a group, $S = S_1 \cup S_2$. We have*

$$|S| + |\operatorname{stab} S| \geq \min\left(|S_i| + |\operatorname{stab} S_i|\right). \tag{3.3}$$

PROOF. The claim is obvious if $S_i = S$ for either i, so we assume they are proper (and consequently nonempty) subsets.

Write $\operatorname{stab} S_i = H_i$, $\operatorname{stab} S = H$. We may also assume that $H_0 = H_1 \cap H_2 = \{0\}$, since otherwise every set is a union of cosets of H_0 and the claim can be reduced to the corresponding claim in the factor group G/H_0.

Write $|H_i| = h_i$. Let $\overline{H} = H_1 + H_2$; clearly $\left|\overline{H}\right| = h_1 h_2$. Each coset of $\overline{H}$ is the union of h_2 cosets of H_1 as well as h_1 cosets of H_2.

We can rewrite (3.3) as

$$|S \setminus S_i| \geq h_i - |H| \text{ for some } i. \tag{3.4}$$

We shall see how to find lower estimates for $|S \setminus S_i|$.

Consider a typical nonempty intersection of S with a coset of $\overline{H}$, say $\overline{H} + x$. Some of the h_2 cosets of H_1 inside it, say k_1, are in S_1, and some k_2 of the h_1 cosets of H_2 are in S_2. From each coset of H_1 exactly k_2 elements are in S_2 and $h_1 - k_2$ in $S \setminus S_2$, hence

$$\left|(S \setminus S_2) \cap (\overline{H} + x)\right| = k_1(h_1 - k_2)$$

and similarly

$$\left|(S \setminus S_1) \cap (\overline{H} + x)\right| = k_2(h_2 - k_1).$$

If there is a coset of $\overline{H}$ such that $0 < k_1 < h_2$ and $0 < k_2 < h_1$, then we multiply the above equations to obtain

$$|S \setminus S_2|\,|S \setminus S_1| \geq k_1 k_2 (h_1 - k_2)(h_2 - k_1) \geq (h_1 - 1)(h_2 - 1),$$

hence at least one of the inequalities

$$|S \setminus S_i| \geq h_i - 1$$

is true and we are done.

If there is no such coset, but there is one in which $k_1 = 0 < k_2$ and a different one in which $k_2 = 0 < k_1$, then by using the first coset to estimate $S \setminus S_1$ and the second to estimate $S \setminus S_2$ we get

$$|S \setminus S_2|\,|S \setminus S_1| \geq h_1 h_2,$$

stronger than before.

Finally assume that one of the above possibilities is missing, say the first. In this case we claim that S is a union of cosets of H_1. We check this on each coset of $\overline{H} + x$. This happens obviously if $\overline{H} + x \subset S$. If this inclusion fails, then clearly $k_1 < h_2$ and $k_2 < h_1$, so one of them must vanish; we excluded $k_1 = 0 < k_2$, so $k_2 = 0$.

This means $H \supset H_1$ and then for $i = 1$ the right side of (3.4) is ≤ 0. □

LEMMA 3.4. *Let S be a finite set in a group, $S = S_1 \cup S_2 \cup \cdots \cup S_k$. We have*

$$|S| + |\operatorname{stab} S| \geq \min \left(|S_i| + |\operatorname{stab} S_i| \right). \tag{3.5}$$

This follows from the previous one by an immediate induction.

PROOF OF KNESER'S THEOREM. Fix a $b \in B$, and consider all possible finite sets $A_b, B_b \subset G$ with the properties

$$b \in B_b,\ A_b \supset A,\ A_b + B_b \subset A + B, |A_b| + |B_b| = |A| + |B|\,. \tag{3.6}$$

Such sets do exist, for instance, $A_b = A$, $B_b = B$. Fix from among them one for which $|B_b|$ is minimal. Put $S_b = A_b + B_b$. We have

$$\bigcup S_b = S.$$

Indeed, one inclusion follows from the second inclusion in (3.6), the other from $A_b + b \subset S_b$.

We try to find a pair with smaller $|B_b|$ by a transfusion:

$$B' = B_b \cap (A_b - t),\ A' = A_b \cup (B_b + t).$$

To preserve the first condition in (3.6) we need $b \in A_b - t$, that is,

$$t \in A_b - b.$$

The inclusions and the equality of cardinality sums hold automatically. The minimality assumption means that each such B' satisfies $B' = B_b$, that is, $B_b \subset A_b - t$, hence $B_b + t \subset A_b$. Forming the union of these inclusions we obtain

$$A_b \supset \bigcup_{t \in A_b - b} (B_b + t) = A_b + B_b - b.$$

This can be reformulated as

$$B_b - b \subset \operatorname{stab} A_b.$$

Clearly $\operatorname{stab} S_b \supset \operatorname{stab} A_b$, so for each b we have

$$|S_b| + |\operatorname{stab} S_b| \geq |S_b| + |B_b| \geq |A_b| + |B_b| = |A| + |B|$$

An application of the previous lemma to these sets S_b gives

$$|S| + |\operatorname{stab} S| \geq \min \left(|S_b| + |\operatorname{stab} S_b| \right) \geq |A| + |B|\,.$$

If we apply this inequality to the sets $A + H$ and $B + H$, since $A + H + B + H = A + B$ and $stab(A + H + B + H) = H$, we obtain (3.1). To get inequality (3.2) observe that each quantity in (3.1) is a multiple of $|H|$, so if they are not equal, then the left exceeds the right at least by $|H|$. □

EXERCISE 61. What is the size of the largest sumfree subset of $\mathbb{Z}_n$, n composite?

4. Sumsets and diameter, part 1

In this section we prove Freiman's Theorem 1.3 in a generalized form.

Translate A so that its minimal element is 0, and divide each element by their greatest common divisor. After these operations we can write A as

$$A = \{a_1, \dots, a_m\},\ a_1 = 0,\ a_m = u$$

and we know that

$$\gcd(a_1, \dots, a_m) = 1.$$

Under these conditions the claim is $|2A| \geq \min(m + u, 3m - 3)$.

This theorem can be extended to the addition of different sets in several ways; we mention two possibilities. In both let $A, B \subset \mathbb{Z}$, $A = \{a_1, \dots, a_m\}$, $B = \{b_1, \dots, b_n\}$ with $0 = a_1 < \cdots < a_m = u$, $0 = b_1 < \cdots < b_n = v$.

THEOREM 4.1 (Freiman [**9**]). *If* $\gcd(a_1, \dots, a_m, b_1, \dots, b_n) = 1$ *and* $u \leq v$, *then*

$$|A + B| \geq \min(m + v, m + n + \min(m, n) - 3).$$

THEOREM 4.2 (Lev and Smelianski [**30**]). *If* $\gcd(b_1, \dots, b_n) = 1$ *and* $u \leq v$, *then*

$$|A + B| \geq \min(m + v, n + 2m - 2 - \delta), \tag{4.1}$$

where $\delta = 1$ *if* $u = v$ *and* $\delta = 0$ *if* $u < v$.

PROOF. Let A', B' be the images of A, B in $\mathbb{Z}_v$; we have

$$|A'| = m' = m - \delta,\ |B'| = n' = n - 1.$$

Kneser's theorem tells us

$$|A' + B'| \geq |A' + H| + |B' + H| - |H| \tag{4.2}$$

with $H = \mathrm{stab}(A' + B')$. Write $|H| = q$. We have $q|v$ (and then H consists exactly of the multiples of v/q). The choice of the two possibilities in (4.1) depends on whether $q = v$ or $q < v$.

In any case we have

$$|A + B| \geq |A' + B'| + m. \tag{4.3}$$

Indeed, $A + B$ has at least one element in each residue class of $A' + B'$. We can exhibit m classes when it has at least two, namely those of $a_1, \dots, a_m$ where a_i and $a_i + v$ are those elements if $u < v$. If $u = v$, this is only $m - 1$ classes, but in the class of 0 there are 3 elements, 0, v and $2v$. If $A' + B' = \mathbb{Z}_v$, this gives us the required $v + m$.

If H is a proper subgroup, we will improve (4.3) as follows. Write

$$|A' + H| = kq,\ |B' + H| = lq.$$

We have $l \geq 2$; indeed, B cannot be in a proper subgroup by the assumption

$$\gcd(b_1, \dots, b_n) = 1.$$

The set $A' + B'$ consists of $\geq k + l - 1 > k$ cosets, so there is one free of elements of A'. Fix such a coset. If $A + B$ has t elements with residues in this coset, we can improve (4.3) to

$$|A + B| \geq |A' + B'| + m + (t - q), \tag{4.4}$$

since in (4.3) only the excess in classes of A was counted.

This coset is the sum of a coset in $A' + H$ and one in $B' + H$. Assume that A and B have r and s elements with residues in these classes, respectively. Then $A + B$ has at least $r + s - 1$, so (4.4) implies

$$|A + B| \geq |A' + B'| + m + r + s - 1 - q. \tag{4.5}$$

In these classes A', B' have at most r and s elements, while $A' + H$, $B' + H$ have exactly q, so we have

$$|A' + H| \geq m' + q - r, \ |B' + H| \geq n' + q - s.$$

On substituting this into (4.2) and applying (4.5) we obtain

$$|A + B| \geq m' + n' + m - 1 = 2m + n - 2 - \delta.$$

□

5. The impact function

Let G be a semigroup (in most cases it will be a commutative group).

DEFINITION 5.1. For a fixed finite set $B \subset G$ we define its *impact function* by

$$\xi_B(m) = \xi_B(m, G) = \min\{|A + B| : A \subset G, |A| = m\}.$$

This is defined for all positive integers if G is infinite, and for $m \leq |G|$ if G is finite.

This function embodies what can be told about cardinality of sumsets if one of the set is unrestricted up to cardinality. The name is a translation of Plünnecke's "Wirkungsfunktion", who first studied this concept systematically for density [**38**].

Some of the previous results, like the Cauchy-Davenport inequality, can be reformulated with this concept; some, like Lev and Smeliansky's Theorem 4.2 cannot, since about A other assumptions than its size are also used.

EXERCISE 62. Let G be a finite group. Prove the following "sort of concavity" of the impact function: for $2 \leq n < |G|$, $n \nmid |G|$ there is a number $1 \leq k \leq n-1$ such that

$$\xi(n - k) + \xi(n + k) \leq 2\xi(n).$$

EXERCISE 63. Use the previous exercise to deduce the Cauchy-Davenport inequality.

EXERCISE 64. (= Exercise 9). Let A, B be finite sets in a (not necessarily commutative) torsionfree group. Show that

$$|A + B| \geq |A| + |B| - 1.$$

EXERCISE 65. In a finite group the graph of the impact function has a certain symmetry with respect to the line $x + y = |G|$. Formulate exactly and prove.

We show that the dependence on G can be omitted.

THEOREM 5.2. *Let G' be a commutative group, G a subgroup of G', and let $B \subset G$ be a finite set. If G is infinite, we have*

$$\xi_B(m, G') = \xi_B(m, G) \tag{5.1}$$

for all m. If G is finite, say $|G| = q$, then for $m = kq + r$, $0 \leq r \leq q - 1$ we have

$$\xi_B(m, G') = \xi_B(r, G) + kq. \tag{5.2}$$

PROOF. Take an $A \subset G'$, $|A| = m$ with $|A+B| = \xi_B(m, G')$. Let $A = A_1 \cup \cdots \cup A_k$ be its decomposition according to cosets of G. For each $1 \le i \le k$ take an element x_i from the coset containing A_i so that the sets $A_i - x_i$ are pairwise disjoint; this is easily done as long as G is infinite. The set

$$A' = \bigcup (A_i - x_i)$$

satisfies $A' \subset G$, $|A'| = m$ and

$$|A' + B| \le \sum |A_i - x_i + B| = \sum |A_i + B| = |A + B| = \xi_B(m, G'),$$

hence $\xi_B(m, G) \le \xi_B(m, G')$. The inequality in the other direction is obvious.

In the finite case from all the sets A at which the minimum is attained select one for which k is minimal, and with k so fixed $\min |A_i|$ is minimal. We claim that all but one A_i are cosets of G; this clearly implies (5.2).

Assume $|A_1| \le \cdots \le |A_k|$ and $A_i \subset G + x_i$. We try to replace A_1, A_2 by sets

$$A_1' = A_1 \cap (A_2 - y),\ A_2' = (A_1 + y) \cup A_2$$

with suitable $y \in x_2 - x_1 + G$.

We claim that this operation does not change the cardinality of A and does not increase that of $A + B$. Indeed,

$$|A_1'| + |A_2'| = |A_1' + y| + |A_2'| = |(A_1 + y) \cap A_2| + |(A_1 + y) \cup A_2| =$$

$$|A_1 + y| + |A_2| = |A_1| + |A_2|\,.$$

Write $A_i + B = C_i$, $A_i' + B = C_i'$. Then

$$C_1' \subset C_1 \cap (C_2 - y),\ C_2' = (C_1 + y) \cup C_2$$

and the comparison of cardinalities goes like for A_1, A_2.

If A_2 does not fill the complete coset, we can find y so that a prescribed element of A_1 be missing from A_1' which would give an example with smaller $|A_1|$, or smaller k if $A_1' = \emptyset$. □

This proof was adapted from arguments in chapters 3 and 4 of Plünnecke's above mentioned book [**38**].

In view of this result we will omit the ambient group G from the notation and write just $\xi_B(m)$ instead.

Let G be a torsionfree group. Take a finite $B \subset G$, and let G' be the subgroup generated by $B - B$, that is, the smallest subgroup such that B is contained in a single coset. Let $B' = B - a$ with some $a \in B$, so that $B' \subset G'$. The group G', as any finitely generated torsionfree group, is isomorphic to the additive group $\mathbb{Z}^d$ for some d. Let $\varphi : G' \to \mathbb{Z}^d$ be such an isomorphism and $B'' = \varphi(B')$. By Theorem 5.2 we have

$$\xi_B = \xi_{B'} = \xi_{B''},$$

so when studying the impact function we can restrict our attention to sets in $\mathbb{Z}^d$ that contain the origin and generate the whole lattice; we then study the set "in its natural habitat".

DEFINITION 5.3. Let B be a finite set in a torsionfree group G. By the *dimension* of B we mean the number d defined above, and denote it by $\dim B$.

Observe that this dimension is not necessarily equal to the geometrical dimension. In the case when $B \subset \mathbb{R}^k$ with some k, this is its dimension over the field of rationals.

The reduced diameter makes sense exactly for one-dimensional sets.

6. Estimates for the impact function in one dimension

We give some estimates that use the diameter and cardinality.

It is possible to give an estimate using the diameter only.

THEOREM 6.1. *Let B be a one-dimensional set in a torsionfree commutative group,* $\operatorname{diam} B = v \geq 3$.
(a) For

$$m > \frac{(v-1)(v-2)}{2}$$

we have $\xi_B(m) = m + v$.
(b) If

$$\frac{(k-1)(k-2)}{2} < m \leq \frac{k(k-1)}{2}$$

with some integer $2 \leq k < v$, then $\xi_B(m) \geq m + k$.

Equality holds for the set $B = \{0, 1, v\} \subset \mathbb{Z}$.

For $v \leq 2$ we have obviously $\xi_B(m) = m+v$ for all m (such a set cannot be anything else than a $v+1$-term arithmetic progression).

This will be deduced from the following result, where the cardinality of B is also taken into account.

THEOREM 6.2. *Let B be a one-dimensional set in a torsionfree commutative group,* $\operatorname{diam} B = v \geq 3$, $|B| = n$. *Define w by*

$$w = \min_{d|v,\, d \leq n-2} d \left[\frac{n-2}{d} \right]. \tag{6.1}$$

For every m we have

$$\xi_B(m) \geq m + \min\left(v, \frac{w}{2} + \min_{t \in \mathbb{N}} \left(\frac{m}{t} + \frac{tw}{2} \right) \right). \tag{6.2}$$

The minimum is attained at one of the integers surrounding $\sqrt{2m/w}$. Unlike the previous theorem, typically we don't have examples of equality, and the extremal value and the structure of extremal sets is probably complicated. Also the value of w depends on divisibility properties of v and n. After the proof we give some less exact but simpler corollaries.

PROOF. By Lemma 5.2 we may assume that $B \subset \mathbb{Z}$, its smallest element is 0 and it generates $\mathbb{Z}$; then its largest element is just v.

LEMMA 6.3. *Let B' be the set of residues of elements of B modulo v. For every nonempty $X \subset \mathbb{Z}_v$ we have*

$$|X + B'| \geq \min(|X| + w, v). \tag{6.3}$$

PROOF. By Kneser's theorem we have

$$|X + B'| \geq |X + H| + |B' + H| - |H|$$

with some subgroup H of the additive group $\mathbb{Z}_v$. Write $|H| = d$; clearly $d|v$. If $d = v$, we have $|X + H| = v$ and we are ready. Assume $d < v$. B' contains 0 and it generates $\mathbb{Z}_v$, hence it cannot be contained in H so we have $|B' + H| \geq 2|H| = 2d$. This gives the desired bound if $d > n - 2$. Assume $d \leq n - 2$. Since $|B' + H|$ is a multiple of d and it is at least $|B'| = n - 1$, we obtain

$$|B' + H| \geq d \left\lceil \frac{n-1}{d} \right\rceil = d \left(1 + \left\lceil \frac{n-2}{d} \right\rceil \right) \geq d + w.$$

□

We resume the proof of Theorem 6.2. Take a set $A \subset \mathbb{Z}$, $|A| = m$. We are going to estimate $|A + B|$ from below.

For $j \in \mathbb{Z}_v$ let $u(j)$ be the number of integers $a \in A$, $a \equiv j \pmod{v}$ and let $U(j)$ be the corresponding number for the sumset $A + B$. We have

$$U(j) \geq u(j) + 1 \tag{6.4}$$

whenever $U(j) > 0$; this follows by adding the numbers $0, v$ to each element of A in this residue class if $u(j) > 0$, and holds obviously for $u(j) = 0$. We also have

$$U(j) \geq u(j - b) \tag{6.5}$$

for every $b \in B'$. Write

$$r(k) = \{j : u(j) \geq k\},$$
$$R(k) = \{j : U(j) \geq k\}.$$

Inequality (6.4) implies

$$R(k) \supset r(k - 1) \ (k \geq 2), \tag{6.6}$$

and inequality (6.5) implies

$$R(k) \supset r(k) + B' \ (k \geq 1). \tag{6.7}$$

First case: $U(j) > 0$ for all j. In this case by summing (6.4) we get

$$|A + B| = \sum U(j) \geq v + \sum u(j) = |A| + v.$$

Second case: there is a j with $U(j) = 0$. Then we have $|R(k)| < v$ for every $k > 0$. An application of Lemma 6.3 to the sets $r(k)$ yields, by view of (6.7)

$$|R(k)| \geq |r(k)| + w \tag{6.8}$$

as long as $r(k) \neq \emptyset$. Let t be the largest integer with $r(t) > 0$. We have (6.8) for $1 \leq k \leq t$, and (6.6) yields

$$|R(k)| \geq |r(k - 1)| \tag{6.9}$$

for all $k \geq 2$. Consequently for $1 \leq k \leq t + 1$ we have

$$|R(k)| \geq \frac{k-1}{t} |r(k-1)| + \left(1 - \frac{k-1}{t} \right) (|r(k)| + w). \tag{6.10}$$

Indeed, for $k = 1$ (6.10) is identical with (6.8), for $k = t + 1$ it is identical with (6.9) and for $2 \leq k \leq t$ it is a linear combination of the two.

By summing (6.10) we obtain

$$|A+B| = \sum_{k\geq 1} |R(k)| \geq \sum_{k=1}^{t+1} |R(k)|$$

$$\geq \frac{t+1}{2}w + \left(1+\frac{1}{t}\right)\sum_{k=1}^{t} |r(k)| = \frac{t+1}{2}w + \left(1+\frac{1}{t}\right)|A|,$$

as claimed in (6.2).

□

COROLLARY 6.4. *With the assumptions and notations of Theorem 6.2 we have*

$$\xi_B(m) \geq \min\left(m+v, \left(\sqrt{m}+\sqrt{w/2}\right)^2\right). \tag{6.11}$$

PROOF. This follows from (6.2) and the inequality of arithmetic and geometric means. □

PROOF OF THEOREM 6.1. Parts (a)-(b) of the theorem can be reformulated as follows: if $\xi_B(m) \leq m+k$ with some $k < v$, then $m \leq k(k-1)/2$. Theorem 6.2 yields (using only that $w \geq 1$) the existence of a positive integer t such that

$$\frac{m}{t} + \frac{t+1}{2} \leq k,$$

hence

$$m \leq kt - \frac{t(t+1)}{2}.$$

The right side, as a function of t, is increasing up to $k-1/2$ and decreasing afterwards; the maximal values at integers are assumed at $t = k-1$ and k, and both are equal to $k(k-1)/2$.

To show the case of equality in case (b), write $m = k(k-1)/2 - l$ with $0 \leq l \leq k-2$. The set A will contain the integers in the intervals $[iv, iv+k-3-i]$ for $0 \leq i \leq l-1$ and $[iv, iv+k-2-i]$ for $l \leq i \leq k-2$.

□

In comparison to the results of Section 4 observe that they never give an increment exceeding $2n$, they are, however, better for small values of m.

PROBLEM 6.5. Find a common generalization of Theorems 6.2 and 4.2.

7. Multidimensional sets

The first result that connects additive properties to geometrical dimension is perhaps the following theorem of Freiman.

THEOREM 7.1 (Freiman[**10**], Lemma 1.14). *Let $A \subset \mathbb{R}^d$ be a finite set, $|A| = m$. Assume that A is proper d-dimensional, that is, it is not contained in any affine hyperplane. Then*

$$|A+A| \geq (d+1)m - \frac{d(d+1)}{2}.$$

PROOF. We use induction on m.

The starting case is $m = 2$. Then necessarily $d = 1$ and the claim is $|2A| \geq 3$, which is indeed true.

Assume now that the statement is true for m and all possible values of d (which are $1 \leq d \leq m-1$). We prove it for $m+1$. Let $A \subset \mathbb{R}^d$, $|A| = m+1$. Consider the convex hull of A, and let $a \in A$ be one of its vertices. Put $A' = A \setminus \{a\}$. We have $|A'| = m$, so the statement is true for A'.

The dimension of A' may be d or $d-1$.

Assume first that this dimension is d. Consider the supporting planes of the conves hull A'. Such a plane L intersects conv A' in one of its sides, whose vertices are elements of A', hence $|A' \cap L| \geq d$, and the rest of A' is on one side. At least one of these planes has the property that a is on the other side. Fix such a plane. Then none of the points of $a + (A' \cap L)$ is a point of $2A'$. Hence

$$|2A| \geq |2A'| + |A' \cap L| + 1$$

(the $+1$ comes from the element $2a$)

$$\geq (d+1)m - \frac{d(d+1)}{2} + d + 1 = (d+1)(m+1) - \frac{d(d+1)}{2} + d + 1$$

as wanted.

Consider now the case when the dimension of A' is $d-1$. Then A' lies on a plane L and the point a is outside it, hence the sets $2A'$, $a + A'$ and $\{2a\}$ are all disjoint and we obtain

$$|2A| \geq |2A'| + |A'| + 1$$

$$\geq dm - \frac{(d-1)d}{2} + m + 1 = (d+1)(m+1) - \frac{d(d+1)}{2} + d + 1$$

again. □

This theorem is exact, equality can occur, namely it holds when A is a "long simplex", a set of the form

$$L_{dm} = \{0, e_1, 2e_1, \ldots, (m-d)e_1, e_2, e_3, \ldots, e_d\}. \tag{7.1}$$

In particular, if no assumption is made on the dimension, then the minimal possible cardinality of the sumset is $2m-1$, with equality for arithmetic progressions.

This result can be extended to sums of different sets. This extension is problematic from the beginning, namely the assumption "d-dimensional" can be interpreted in different ways. We can stipulate that both sets be d-dimensional, or only one, or, in the weakest form, make this assumption on the sumset only.

An immediate extension of Freiman's above result goes as follows.

THEOREM 7.2 ([**54**], Corollary 1.1). *If $A, B \subset \mathbb{R}^d$, $|A| \leq |B|$ and $\dim(A+B) = d$, then we have*

$$|A+B| \geq |B| + d|A| - \frac{d(d+1)}{2}.$$

We can compare these results to to the continuous case. Let A, B be Borel sets in $\mathbb{R}^d$; μ will denote the Lebesgue measure. The celebrated Brunn-Minkowski inequality asserts that

$$\mu(A+B)^{1/d} \geq \mu(A)^{1/d} + \mu(B)^{1/d}, \tag{7.2}$$

and here equality holds if A and B are homothetic convex sets, and under mild and natural assumptions this is the only case of equality. It can also be observed that the case $A = B$ is completely obvious here: we have

$$\mu(A + A) \geq \mu(2 \cdot A) = 2^d \mu(A).$$

Also the constant 2^d is much larger than the constant $d + 1$ in Theorem 7.1. This is necessary, as there are examples of equality, however, one feels that this is an exceptional phenomenon and better estimations should hold for "typical" sets. A further difference is the asymmetrical nature of the discrete result and the symmetry of the continuous one. Finally, when $|A|$ is fixed, Theorem 7.2 gives a linear increment, while (7.2) yields

$$\mu(A + B) \geq \mu(B) + d\mu(A)^{1/d}\mu(B)^{1-1/d}.$$

In the next section we tell what can be said if we use cardinality as the discrete analog of measure, and prescribe only the dimension of the sets. Later we try to find other spatial properties that may be used to study sumsets.

The main problems are perhaps the following. What are the best analogs of measure and dimension for discrete sets? How should a discrete analog of the Brunn-Minkowski inequality look like? The partial answers explained below also suggest questions in the continuous case. Should we be satisfied with the usual concepts of measure and dimension for studying the addition of sets? We return to this in Chapter 5.

Mostly our sets will be in an Euclidean space $\mathbb{R}^d$, and $e_1, \ldots, e_d$ will be the system of unit vectors. We can think of the *dimension* $\dim A$ of a set $A \subset \mathbb{R}^d$ as the dimension of the smallest affine hyperplane containing A, or as in Definition 5.3.

8. Results using cardinality and dimension

We consider finite sets in an Euclidean space $\mathbb{R}^d$.

Put

$$F_d(m, n) = \min\{|A + B| : |A| = m, |B| = n,\ \dim(A + B) = d\}$$

$$F'_d(m, n) = \min\{|A + B| : |A| = m, |B| = n,\ \dim B = d\},$$

$$F''_d(m, n) = \min\{|A + B| : |A| = m, |B| = n,\ \dim A = \dim B = d\}.$$

F_d is defined for $m + n \geq d + 2$, F'_d for $n \geq d + 1$ and F''_d for $m \geq d + 1, n \geq d + 1$. F_d and F''_d are obviously symmetric, while F'_d may not be (and, in fact, we will see that for certain values of m, n it is not), and they are connected by the obvious inequalities

$$F_d(m, n) \leq F'_d(m, n) \leq F''_d(m, n).$$

I determined the behaviour of F_d and of F'_d for $m \leq n$. The more difficult problem of describing F''_d and F'_d for $m > n$ was solved by Gardner and Gronchi [**12**]; we shall quote their results later.

To describe F_d define another function G_d as follows:

$$G_d(m, n) = n + \sum_{j=1}^{m-1} \min(d, n - j), \quad n \geq m \geq 1$$

and for $m > n$ extend it symmetrically, putting $G_d(m, n) = G_d(n, m)$. In other words, if $n - m \geq d$, then we have

$$G_d(m, n) = n + d(m - 1).$$

If $0 \le t = n - m < d$, then for $n > d$ we have

$$G_d(m,n) = n + d(m-1) - \frac{(d-t)(d-t-1)}{2} = n(d+1) - \frac{d(d+1)}{2} - \frac{t(t+1)}{2},$$

and for $n \le d$

$$G_d(m,n) = n + \frac{(m-1)(2n-m)}{2}.$$

With this notation we have the following result.

THEOREM 8.1 ([**54**], Theorem 1). *For all positive integers m, n and d satisfying $m + n \ge d + 2$ we have*

$$F_d(m,n) \ge G_d(m,n).$$

Theorem 7.2 is an immediate consequence.

Theorem 8.1 is typically exact; the next theorem summarizes the cases when we have examples of equality.

THEOREM 8.2 ([**54**], Theorem 2). *Assume $1 \le m \le n$. We have*

$$F_d(m,n) = F'_d(m,n) = G_d(m,n)$$

unless either $n < d+1$ or $m \le n - m \le d$ (in this case $n \le 2d$).

The construction goes as follows.

Assume $1 \le m \le n$, $n \ge d+1$. Let B be a long simplex, $B = L_{dn}$ as defined in (7.1).

If $n - m \ge d$, we put

$$A = \{0e_1, 1e_1, \dots, (m-1)e_1\}.$$

This set satisfies $|A| = m$. The set $A+B$ consists of the vectors ie_1, $0 \le i \le n+m-d-1$ and the vectors $ie_1 + e_j$, $0 \le i \le m-1$, $2 \le j \le d$, consequently

$$|A + B| = n + d(m-1) = G_d(m,n).$$

If $n - m = t < d$, write $t = d - k$ and assume $k \le m$. Now A is defined by

$$A = \{0e_1, 1e_1, \dots, (m-k)e_1\} \cup \{e_2, \dots, e_k\}.$$

This set satisfies $|A| = m$. The set $A + B$ consists of the vectors ie_1, $0 \le i \le 2(n-d)$, the vectors $ie_1 + e_j$, $0 \le i \le n-d$, $2 \le j \le d$, finally $e_i + e_j$, $2 \le i \le k$, $2 \le j \le d$, hence

$$|A + B| = 2(n-d) + 1 + (d-1)(n-d+1) + d(k-1) - \frac{k(k-1)}{2}$$

$$= n(d+1) - \frac{d(d+1)}{2} - \frac{t(t+1)}{2} = G_d(m,n).$$

These constructions cover all pairs m, n except those listed in Theorem 8.2. Observe that A is also a long simplex of lower dimension. For a few small values the exact bounds are yet to be determined.

We now describe Gardner and Gronchi's [**12**] bound for $F'_d(m,n)$. Informally their main result (Theorem 5.1) asserts that the $|A+B|$ is minimalized when $B = L_{dn}$, a long simplex, and A is as near to the set of points inside a homothetic simplex as possible. More exactly they define (for a fixed value of n) the weight of a point $x = (x_1, \dots, x_d)$ as

$$w(x) = \frac{x_1}{n-d} + x_2 + \cdots + x_d.$$

This defines an ordering by writing $x < y$ if either $w(x) < w(y)$ or $w(x) = w(y)$ and for some j we have $x_j > y_j$ and $x_i = y_i$ for $i < j$.

Let D_{dmn} be the collection of the first m vectors with nonnegative integer coordinates in this ordering. We have $D_{dnn} = L_{dn} = B$, and, more generally, $D_{dmn} = rB$ for any integer m such that

$$m = |rB| = (n-d)\binom{r+d-1}{d} + \binom{r+d-1}{d-1}.$$

For such values of m we also have

$$|A+B| = |(r+1)B| = (n-d)\binom{r+d}{d} + \binom{r+d}{d-1}.$$

With this notation their result sounds as follows.

THEOREM 8.3 (Gardner and Gronchi[**12**], Theorem 5.1). *If $A, B \subset \mathbb{R}^d$, $|A| = m$, $|B| = n$ and $\dim B = d$, then we have*

$$|A+B| \geq |D_{dmn} + L_{dn}|.$$

For $m < n$ this reproves Theorem 8.2. For $m \geq n$ the extremal set D_{dmn} is also d-dimensional, thus this result also gives the value of F''_d.

COROLLARY 8.4. *For $m \geq n > d$ we have*

$$F''_d(m,n) = F'_d(m,n) = |D_{dmn} + L_{dn}|.$$

A formula for the value of this function is given in [**12**], Section 6. We quote some interesting consequences.

THEOREM 8.5 (Gardner and Gronchi[**12**], Theorem 6.5). *If $A, B \subset \mathbb{R}^d$, $|A| = m \geq |B| = n$ and $\dim B = d$, then we have*

$$|A+B| \geq m + (d-1)n + (n-d)^{1-1/d}(m-d)^{1/d} - \frac{d(d-1)}{2}.$$

THEOREM 8.6 (Gardner and Gronchi[**12**], Theorem 6.6). *If $A, B \subset \mathbb{R}^d$, $|A| = m$, $|B| = n$ and $\dim B = d$, then we have*

$$|A+B|^{1/d} \geq m^{1/d} + \left(\frac{n-d}{d!}\right)^{1/d}.$$

This result is as close to the Brunn-Minkowski inequality as we can get by using only the cardinality of the summands.

9. The impact function and the hull volume

Let G be a torsionfree group. Take a finite $B \subset G$. At the end of Section 5 we defined a certain "natural image" of B as follows. Let G' be the subgroup generated by $B - B$ and $B' = B - a$ with some $a \in B$, so that $B' \subset G'$. The group G' is isomorphic to the additive group $\mathbb{Z}^d$ for some d. Let $\varphi : G' \to \mathbb{Z}^d$ be such an isomorphism and $B'' = \varphi(B')$. By Theorem 5.2 we know

$$\xi_B = \xi_{B'} = \xi_{B''}.$$

so when studying the impact function we can restrict our attention to sets in $\mathbb{Z}^d$ that contain the origin and generate the whole lattice. We used this d as a definition for an "intrinsic dimension". This image has further usages.

DEFINITION 9.1. Let B be a finite set in a torsionfree group G. By the *hull volume* of B we mean the volume of the convex hull of the set B'' described above and denote it by $\operatorname{hv} B$.

The set B'' is determined up to an automorphism of $\mathbb{Z}^d$. These automorphisms are exactly linear maps of determinant ± 1, hence the hull volume is uniquely defined.

THEOREM 9.2. *Let B be a finite set in a torsionfree group G, $d = \dim B$, $v = \operatorname{hv} B$. We have*

$$\lim |kB| k^{-d} = v.$$

A proof can be found in [**55**], Section 11, though this form is not explicitly stated there. An outline is as follows. By using the arguments above we may assume that $B \subset \mathbb{Z}^d$, $0 \in B$ and B generates $\mathbb{Z}^d$. Let B^* be the convex hull of B. Then kB is contained in $k \cdot B^*$. The number of lattice points in $k \cdot B$ is asymptotically $\mu(k \cdot B^*) = k^d v$; this yields an upper estimate. To get a lower estimate one proves that with some constant p, kB contains all the lattice points inside translate of $(k-p) \cdot B^*$; this is Lemma 11.2 of [**55**].

This means that the hull volume can be defined without any reference to convexity and measure. Later we will show that this definition can even be extended to commutative semigroups.

It turns out that in $\mathbb{Z}^d$, hence in any torsionfree group, the dimension and hull volume determine the asymptotic behaviour of the impact function.

THEOREM 9.3. *Let B be a finite set in a torsionfree commutative group G, $d = \dim B$, $v = \operatorname{hv} B$. We have*

$$\lim \xi_B(m)^{1/d} - m^{1/d} = v^{1/d}.$$

This is the main result (Theorem 3.1) of [**55**]. In the same paper I announce the same result for non necessarily torsionfree commutative groups without proof (Theorem 3.4). In a general semigroup $A + B$ may consist of a single element, so an attempt to an immediate generalization fails.

PROBLEM 9.4. Does the limit $\lim \xi_B(m)^{1/d} - m^{1/d}$ exist in general commutative semigroups? Is there a condition weaker than cancellativity to guarantee its positivity?

Theorem 9.3 can be effectivized as follows (Theorems 3.2 and 3.3 of [**55**]).

THEOREM 9.5. *With the notations of the previous theorem, if $d \geq 2$ and $m \geq v$, we have*

$$\xi_B(m) \leq m + dv^{1/d} m^{1-1/d} + c_1 v^{2/d} m^{1-2/d},$$
$$\xi_B(m)^{1/d} - m^{1/d} \leq v^{1/d} + c_2 v^{2/d} m^{-1/d}$$

(c_1, c_2 depend on d.) With $n = |B|$ for large m we have

$$\xi_B(m) \geq m + dv^{1/d} m^{1-1/d} - c_3 v^{\frac{d+3}{2d}} n^{-1/2} m^{1-\frac{3}{2d}},$$
$$\xi_B(m)^{1/d} - m^{1/d} \geq v^{1/d} - c_4 v^{\frac{d+3}{2d}} n^{-1/2} m^{-1/(2d)}.$$

Probably the real error terms are much smaller than these estimates. For $d = 1$ we have the obvious inequality $\xi_B(m) \leq m + v$, with equality for large m because the integers $\xi_B(m) - m$ cannot converge to v otherwise. For $d = 2$ already $\sqrt{\xi_B(m)} - \sqrt{m}$ can converge to $\sqrt{v}$ from both directions.

THEOREM 9.6. *The impact function of the set* $B = \{0, e_1, e_2\} \subset \mathbb{Z}^2$ *satisfies*

$$\sqrt{\xi_B(m)} - \sqrt{m} > \sqrt{v} \tag{9.1}$$

for all m.

The impact function of the set $B = \{0, e_1, e_2, -(e_1 + e_2)\} \subset \mathbb{Z}^2$ *satisfies*

$$\sqrt{\xi_B(m)} - \sqrt{m} < \sqrt{v} \tag{9.2}$$

for infinitely many m.

Inequality (9.1) was announced in [**55**] without proof as Theorem 4.1, and it is a special case of Gardner and Gronchi's Theorem 8.6. Inequality 9.2 is Theorem 4.3 of [**55**].

I cannot decide whether there is a set such that $\sqrt{\xi_B(m)} - \sqrt{m} < \sqrt{v}$ for all m.

10. The impact volume

Besides cardinality we saw the hull volume as a contender for the title "discrete volume". For both we had something resembling the Brunn-Minkowski inequality; for cardinality we had Gardner and Gronchi's Theorem 8.6, which has the (necessary) factor $d!$, and for the hull volume we have Theorem 9.3, which only holds asymptotically.

There is an easy way to find a quantity for which the analogue of the Brunn-Minkowski inequality holds exactly: we can make it a definition.

DEFINITION 10.1. The d-dimensional *impact volume* of a set B (in an arbitrarily commutative group) is the quantity

$$\mathrm{iv}_d(B) = \inf_{m \in \mathbb{N}} \left(\xi_B(m)^{1/d} - m^{1/d}\right)^d.$$

Note that the d above may differ from the dimension of B, in fact, it need not be an integer. It seems, however, that the only really interesting case is $d = \dim B$.

The following statement list some immediate consequences of this definition.

STATEMENT 10.2. *Let* B *be a finite set in a commutative torsionfree group.*
(a) $\mathrm{iv}_d(B)$ *is a decreasing function of* d.
(b) If $|B| = n$, *then*

$$\mathrm{iv}_1(B) = n - 1$$

and

$$\mathrm{iv}_d(B) \le \left(n^{1/d} - 1\right)^d \tag{10.1}$$

for every d.
(c) $\mathrm{iv}_d(B) = 0$ *for* $d > dim B$.
(d) For every pair A, B *of finite sets in the same group and every* d *we have*

$$\mathrm{iv}_d(A + B)^{1/d} \ge \mathrm{iv}_d(A)^{1/d} + \mathrm{iv}_d(B)^{1/d}. \tag{10.2}$$

The price we have to pay for the discrete Brunn-Minkowski inequality (10.2) is that there is no easy way to compute the impact volume for a general set. We have the following estimates.

THEOREM 10.3. *Let B be a finite set in a commutative torsionfree group,* $\dim B = d$, $|B| = n$. *We have*

$$\left(\frac{n-d}{d!}\right) \le \mathrm{iv}_d(B) \le \mathrm{hv}\, B, \tag{10.3}$$

with equality in both places if B is a long simplex.

The first inequality follows form Theorem 8.6 of Gardner and Gronchi, the second from Theorem 9.3.

PROBLEM 10.4. What is the *maximal* possible value of $\mathrm{iv}_d(B)$ for n-element d-dimensional sets? Is perhaps the bound in (10.1) exact?

We now describe the impact volume for another important class of sets, namely cubes.

THEOREM 10.5. *Let $n_1, \dots, n_d$ be positive integers and let*

$$B = \{(x_1, \dots, x_d) \in \mathbb{Z}^d : 0 \le x_i \le n_i\}. \tag{10.4}$$

We have

$$\mathrm{iv}_d(B) = \mathrm{hv}\, B = v = n_1 \dots n_d.$$

PROBLEM 10.6. Is is true that when B is the set of lattice points within a convex lattice polytope, then $\mathrm{hv}\, B$ and $\mathrm{iv}_d(B)$ are very near?

They may differ, as the second example in Theorem 9.6 shows.

We shall deduce Theorem 10.5 from the following one.

THEOREM 10.7. *Let $G = G_1 \times G_2$ be a commutative group represented as the direct product of the groups G_1 and G_2. Let $B = B_1 \times B_2 \subset G$ be a finite set with $B_1 \subset G_1$, $B_2 \subset G_2$. We have*

$$\mathrm{iv}_d(B) \ge \mathrm{iv}_{d-1}(B_1)\mathrm{iv}_1(B_2). \tag{10.5}$$

PROOF. Write $\mathrm{iv}_d(B) = v$, $\mathrm{iv}_{d-1}(B_1) = v_1$, $\mathrm{iv}_1(B_2) = v_2$ (which is $= |B_2| - 1$ if G_2 is torsionfree). We want to estimate $|A + B|$ from below for a general set $A \subset G$ with $|A| = m$.

First we transform them to some standard form; this will be the procedure what Gardner and Gronchi call compression. Let A_1 be the projection of A to G_1, and for an $x \in A_1$ write

$$A(x) = \{y \in G_2 : (x, y) \in A\}.$$

Let

$$A' = \{(x, i) : x \in A_1, i \in \mathbb{Z},\ 0 \le i \le |A(x)| - 1\}$$

and

$$B' = \{(x, i) : x \in B_1, i \in \mathbb{Z},\ 0 \le i \le v_2\}.$$

We have $A', B' \subset G' = G_1 \times \mathbb{Z}$.

LEMMA 10.8. *We have*

$$|A'| = |A|,\ |A' + B'| \le |A + B|. \tag{10.6}$$

PROOF. The equality is clear. To prove the inequality, write $S = A+B$, $S' = A'+B'$. With the obvious notation, we will show that

$$|S'(x)| \le |S(x)|$$

for each x. To this end observe that

$$S(x) = \bigcup_{x'+x''=x} (A(x') + B(x'')) = \bigcup_{x' \in x - B_1} A(x') + B_2,$$

hence

$$|S(x)| \ge \max_{x' \in x-B_1} |A(x') + B_2| \ge \max_{x' \in x-B_1} |A(x')| + v_2.$$

Similarly

$$S'(x) = \bigcup_{x'+x''=x} (A'(x') + B'(x'')) = \bigcup_{x' \in x - B_1} [0, |A(x')| + v_2 - 1],$$

and so

$$|S'(x)| = \max_{x' \in x-B_1} |A(x')| + v_2.$$

□

Now we continue the proof of the theorem. Decompose A' into layers according to the value of the second component; write

$$A' = \bigcup_{i=0}^{k} L_i \times \{i\},$$

where $k = \max |A(x)|$, $L_i \subset G_1$. Write $|L_i| = m_i$. We have $L_0 \supset L_1 \supset \cdots \supset L_k$, consequently $m_0 \ge m_1 \ge \cdots \ge m_k$.

The set S' is the union of the sets $(L_i + B_1) \times \{i+j\}$, $0 \le i \le k$, $0 \le j \le v_2$. By the above inclusion it is sufficient to consider the L_i with the smallest possible i, that is,

$$S' = (L_0 + B_1) \times \{0, 1, \dots, v_2\} \cup \bigcup_{i=1}^{k} (L_i + B_1) \times \{i + v_2\}.$$

We obtain that

$$|S'| = v_2 |L_0 + B_1| + \sum_{i=0}^{k} |L_i + B_1|. \tag{10.7}$$

To estimate the summands we use the $d-1$ -dimensional impact of B_1. Recall that by definition this means

$$|X + B_1| \ge \left(|X|^{\frac{1}{d-1}} + v_1^{\frac{1}{d-1}} \right)^{d-1}$$

for any set X. We apply this to the sets L_i to obtain

$$|L_i + B_1| \ge \left(m_i^{\frac{1}{d-1}} + v_1^{\frac{1}{d-1}} \right)^{d-1} \ge \frac{m_i}{m_0} \left(m_0^{\frac{1}{d-1}} + v_1^{\frac{1}{d-1}} \right)^{d-1};$$

the second inequality follows from $m_i \le m_0$. By substituting this into (10.7) and recalling that $\sum m_i = m$ we obtain

$$|S| \ge \left(v_2 + \frac{m}{m_0} \right) \left(m_0^{\frac{1}{d-1}} + v_1^{\frac{1}{d-1}} \right)^{d-1}. \tag{10.8}$$

Consider the right side as a function of the real variable m_0. By differentiating we find that it assumes its minimum at

$$m_0 = v_1^{1/d}(m/v_2)^{1-1/d}.$$

(This minimum typically is not attained; this m_0 may be < 1 or $> m$, and it is generally not integer). Substituting this value of m_0 into (10.8) we obtain the desired bound

$$|S| \geq \left(m^{1/d} + (v_1 v_2)^{1/d}\right)^d.$$

□

PROBLEM 10.9. Does equality always hold in Theorem 10.7?

I expect a negative answer.

PROBLEM 10.10. Can Theorem 10.7 be extended to an inequality of the form

$$\mathrm{iv}_{d_1+d_2}(B_1 \times B_2) \geq \mathrm{iv}_{d_1}(B_1)\mathrm{iv}_{d_2}(B_2)?$$

PROOF OF THEOREM 10.5. To prove $\geq$ we use induction on d. The case $d = 1$ is obvious, and Theorem 10.7 provides the inductive step.

This means that with the cube B defined in (10.4) we have

$$|A + B| \geq \left(|A|^{1/d} + v^{1/d}\right)^d.$$

Equality can occur for infinitely many values of $|A|$, namely it holds whenever A is also a cube of the form

$$A = \{(x_1, \ldots, x_d) \in \mathbb{Z}^d : 0 \leq x_i \leq kn_i - 1\}$$

with some integer k; we have $|A| = k^d v$, $|A + B| = (k + 1)^d v$. It may be difficult to describe $\xi_B(m)$ for values of m which are not of the form $k^d v$. Possibly an argument like Gardner and Gronchi's for the simplex may work.

Observe that these special sets A are not homothetic to B; in particular, $A = B$ may not yield a case of equality. □

As Theorem 10.3 shows, the impact volume can be $d!$ times smaller than cardinality. The example we have of this phenomenon, the long simplex, is, however, "barely" d-dimensional, and I expect that a better estimates hold for a "substantially" d-dimensional set.

DEFINITION 10.11. The *thickness* $\vartheta(B)$ of a set $B \subset \mathbb{R}^d$ is the smallest integer k with the property that there is a hyperplane P of $\mathbb{R}^d$ and $x_1, \ldots, x_k \in \mathbb{R}^d$ such that $B \subset \bigcup_{i=1}^k P + x_j$.

CONJECTURE 10.12. For every $\varepsilon > 0$ and d there is a k such that for every $B \subset \mathbb{R}^d$ with $\vartheta(B) > k$ we have $\mathrm{iv}_d(B) > (1 - \varepsilon)|B|$.

This conjecture would yield a discrete Brunn-Minkowski inequality of the form

$$|A + B|^{1/d} \geq |A|^{1/d} + (1 - \varepsilon)|B|^{1/d}$$

assuming a bound on the thickness of B. Such an inequality is true at least in the special case $A = B$. This can be deduced from a result of Freiman ([**10**], Lemma 2.12; see also Bilu [**2**]), which sounds as follows. If $A \subset \mathbb{R}^d$ and $|2A| < (2^d - \varepsilon)|A|$, then there is a hyperplane P such that $|P \cap A| > \delta|A|$, with $\delta = \delta(d, \varepsilon) > 0$.

11. Hovanskii's theorem

We saw examples that cardinalities of sumsets can behave wildly. We show that, in a rather general setting, if we keep on adding the same set persistently, then these irregularities fade.

THEOREM 11.1. *Let A be a finite set in a commutative semigroup G. There is a polynomial f and an integer n_0 such that for $n > n_0$ we have*

$$|nA| = f(n).$$

This theorem is due to Hovanskii [**24, 25**]. A generalization to the effect that $|n_1A_1 + \cdots + n_kA_k|$ becomes a polynomial if all the n_i are large is given by Nathanson [**34**]. Another proof of this theorem was given by Nathanson and Ruzsa [**35**]. Below we give this proof for the case of one variable only.

Unfortunately there is no way to tell this polynomial (except the leading term, see ...) and the threshold n_0.

PROOF. We assume that G has a zero element. If it does not, extend it by a new element (this is only a notational convenience).

Let $A = \{a_1, \dots, a_m\}$. The elements of nA are all sums of the form

$$b = \sum x_i a_i, \ \sum x_i = n,$$

where the coefficients x_i are nonnegative integers and $0a_i$ is the zero of G. We shall consider these coefficients together in the form of a vector $\mathbf{x} = (x_1, \dots, x_m)$ with nonnegative integer coordinates.

Several vectors may induce the same b. From the possible representations we shall select the *lexicographically first*. We write $\mathbf{x} \prec \mathbf{y}$ and say that $\mathbf{x}$ *precedes* $\mathbf{y}$ if there is an i, $1 \le i \le m$ such that $x_1 = y_1, \dots, x_{i-1} = y_{i-1}$, $x_i < y_i$. By the *rank* $r(\mathbf{x})$ of a vector we mean the sum of its coordinates.

We say that a vector $\mathbf{x}$ is *useless*, if there is a $\mathbf{y} \prec \mathbf{x}$ of the same rank such that $\sum x_i a_i = \sum y_i a_i$, and we call it *useful*, if no such $\mathbf{y}$ exists. With this terminology, $|nA|$ is the number of useful vectors of rank n.

Write $\mathbf{x} \le \mathbf{y}$ if $x_i \le y_i$ for each coordinate, and $\mathbf{x} < \mathbf{y}$ if $\mathbf{x} \le \mathbf{y}$ and $\mathbf{x} \ne \mathbf{y}$. If $\mathbf{x}$ is useless and $\mathbf{x} \le \mathbf{x}'$, then $\mathbf{x}'$ is also useless. Indeed, take a $\mathbf{y}$ of the same rank as $\mathbf{x}$, such that $\sum x_i a_i = \sum y_i a_i$ and $\mathbf{y} \prec \mathbf{x}$. Then by adding $\sum (x_i' - x_i)a_i$ to both sides of this equation we find a vector, namely $\mathbf{y}' = \mathbf{y} + \mathbf{x}' - \mathbf{x}$ that precedes $\mathbf{x}'$, has the same rank and induces the same product.

We say that $\mathbf{z}$ is *primitive useless*, if it is useless and there is no useless $\mathbf{x}$ satisfying $\mathbf{x} < \mathbf{z}$. Clearly a vector $\mathbf{x}$ is useless if and only if there is a primitive useless $\mathbf{z}$ such that $\mathbf{z} \le \mathbf{x}$.

By definition, the primitive useless vectors are all incomparable with respect to the relation $<$. It is a well known (and easy) fact that any collection of incomparable vectors (with nonnegative integer coordinates) must be finite.

EXERCISE 66. Prove this finiteness claim.

Hence there are only finitely many primitive useless vectors, say $\mathbf{z}_1, \dots, \mathbf{z}_k$.

By the sieve formula we have

$$|nA| = \sum_{j=0}^{k} (-1)^j \sum_{i_1,\dots,i_j} B(n; i_1, \dots, i_j)$$

where

$$B(n; i_1, \dots, i_j) = \#\{\mathbf{x} : r(\mathbf{x}) = n, \mathbf{x} \geq \mathbf{z}_{i_1}, \dots, \mathbf{x} \geq \mathbf{z}_{i_j}\}.$$

The system of inequalities $\mathbf{x} \geq \mathbf{z}_{i_t}$ is equivalent to a single inequality $\mathbf{x} \geq \mathbf{z}$, where each coordinate of $\mathbf{z}$ is the maximum of the corresponding coordinates of the vectors $\mathbf{z}_{i_t}$. The number of vectors $\mathbf{x}$ satisfying $\mathbf{x} \geq \mathbf{z}$ and $r(\mathbf{x}) = n$ is 0 if $r(\mathbf{z}) > n$, and it is equal to the number of vectors of rank $n - r(\mathbf{z})$ otherwise. This latter is equal to

$$\binom{n - r(\mathbf{z}) + k - 1}{k - 1},$$

a polynomial in n. Hence all the summands $B(n; \dots)$ are polynomials for large n, thus so is $|nA|$. □

Remark. A corresponding result will not hold without the assumption of commutativity; indeed, if the elements of A generate a free semigroup, then we have $|nA| = k^n$. In a noncommutative semigroup, $|nA|$ need not be monotonically increasing. However, we cannot decide the following.

Problem 11.2. Let S be a noncommutative group. Suppose that there are positive constants c, C such that $|nA| \leq Cn^c$. Does it follow that $|nA|$ is a polynomial for large n?

This theorem enables us to define dimension and volume in semigroups in a way that is compatible with our notions in $\mathbb{Z}^d$.

Definition 11.3. Let B be a finite set in a commutative semigroup, and let vk^d be the leading term of the polynomial which coincides with $|kB|$ for large k. By the *dimension* of B we mean the degree d of this polynomial, and by the *hull volume* we mean the leading coefficient v.

CHAPTER 4

Density

1. Asymptotic and Schnirelmann density

A finite set is naturally measured by its cardinality. A set of reals is naturally measured by its Lebesgue measure (nonmeasurable sets do exist, just we never meet them). There is no similarly universal way to measure and compare infinite sets of integers. The most naturally defined one is the asymptotic density.

For a set A of integers we shall use the same letter to denote its

$$A(x) = |A \cap [1, x]| .$$

We allow A to contain 0 or negative numbers, but they are not taken into account in the counting function.

DEFINITION 1.1. The *asymptotic density* of a set A of integers is defined by

$$d(A) = \lim_{x\to\infty} A(x)/x,$$

if this limit exists. The *lower* and *upper* (asymptotic) densities are the corresponding lower and upper limits, respectively:

$$\underline{d}(A) = \liminf_{x\to\infty} A(x)/x, \quad \overline{d}(A) = \limsup_{x\to\infty} A(x)/x.$$

EXERCISE 67. If $\underline{d}(A) > 0$, is there always an $A' \subset A$ with $d(A') > 0$?

EXERCISE 68. If $\underline{d}(A)+\underline{d}(B) > 1$, then $A+B$ contains all but finitely many positive integers.

EXERCISE 69. Let α, β, γ be positive real numbers such that $\alpha + \beta \le \gamma \le 1$. Construct sets of positive integers such that $d(A) = \alpha$, $d(B) = \beta$, $d(A + B) = \gamma$.

As we mentioned in the introduction, combinatorial additive theory grew out of the classical, by Schnirelmann's approach to the Goldbach problem. Goldbach's conjecture asserts that any integer > 3 can be expressed as a sum of 2 or 3 primes, depending on parity. Schnirelmann proved the weaker result that there is a bound k so that every large enough integer is a sum of at most k primes.

The best universe to work with will be the set $\mathbb{N}_0$ of nonnegative integers.

DEFINITION 1.2. A set $A \subset \mathbb{N}_0$ is an *additive basis of order* h, if $hA = \mathbb{N}_0$, that is, every positive integer can be expressed as a sum of h integers from A.

A set $A \subset \mathbb{N}_0$ is an *asymptotic basis of order* h, if every sufficiently large integer can be expressed as a sum of h integers from A, that is, $\mathbb{N}_0 \setminus hA$ is finite.

The smallest such integer h is called the *exact order* or *exact asymptotic order* of A, respectively.

So the proper wording is that the set P of primes forms an asymptotic basis. To be a basis a set must contain 0 and 1.

To this end Schnirelmann established that integers that can be written as a sum of two primes have positive density; and every set having positive density is a basis. An exact form of the first claim is simply

$$\underline{d}(2P) > 0, \tag{1.1}$$

a result which is (in hindsight) not too difficult to prove by sieve methods. Today we know that almost all even integers can be written as a sum of two primes, hence $d(2P) = 1/2$.

To formulate the second claim exactly Schnirelmann introduced a different notion of density.

DEFINITION 1.3. The *Schnirelmann density* of a set A of integers is the number

$$\sigma(A) = \inf_{n\in\mathbb{N}} A(n)/n.$$

This is a less natural concept than asymptotic density. Asymptotic density is translation invariant and it is invariant under the exclusion or inclusion of finitely many elements; Schnirelmann density does not have either property, in fact, $\sigma(A) = 0$ if $1 \notin A$.

EXERCISE 70. $\sigma(A) > 0$ if and only if $1 \in A$ and $\underline{d}(A) > 0$.

EXERCISE 71. Let $\sigma(A) = \alpha$. Show the existence of an $A' \subset A$ such that $\sigma(A') = \alpha$, but by omitting any single element the density of the remaining set will be $< \alpha$.

EXERCISE 72. Let $B \subset \mathbb{N}_0 = \mathbb{N} \cup \{0\}$ and a number $\alpha \in (0, 1)$ be given, and define β as

$$\beta = \inf\{\sigma(A + B) : \sigma(A) \geq \alpha\}.$$

Show the existence of a set A satisfying $\sigma(A) = \alpha$, $\sigma(A + B) = \beta$.

EXERCISE 73. Show that we have always

$$\underline{d}(A) = \sup \sigma(A - n).$$

Show that we cannot replace the supremum by maximum.

PREXERCISE. If $\sigma(A) + \sigma(B) > 1$ and $0 \in A$, then $A + B \supset \mathbb{N}$. (See Theorem 2.2 below.)

In these terms Schnirelmann's result sounds as follows.

THEOREM 1.4. *If $0 \in A$ and $\sigma(A) > 0$, then A is a basis.*

This theorem will be proved and an estimate for the order of this basis in terms of $\sigma(A)$ will be given in the next sections.

EXERCISE 74. How does it follow from (1.1) and the theorem above that P is an asymptotic basis?

2. Schirelmann's inequality

Schnirelmann deduced his Theorem 1.4 from the following inequality.

THEOREM 2.1. *Let A and B be sets of nonnegative integers with positive Schnirelmann densities $\sigma(A) = \alpha$ and $\sigma(B) = \beta$, respectively. If $0 \in A \cup B$, then*

$$\sigma(A + B) \geq \alpha + \beta - \alpha\beta. \tag{2.1}$$

PROOF. Without restricting generality we can assume $0 \in A$. Put $C = A + B$; we are going to estimate $C(n)$ for an arbitrary positive integer n. Let

$$1 = b_1 < \cdots < b_k \leq n$$

be the elements of B in $[1, n]$. We have $k = B(n) \geq \beta n$. Since $0 \in A$, these numbers are also in C. Further elements of C are given by

$$b_1 + (A \cap [1, b_2 - b_1 - 1]),\ b_2 + (A \cap [1, b_3 - b_2 - 1]),\ \ldots,$$

$$\ldots, b_{k-1} + (A \cap [1, b_k - b_{k-1} - 1]),\ b_k + (A \cap [1, n - b_k]).$$

(The last block may be empty if $b_k = n$.) We estimate the number of elements in a typical block by

$$|A \cap [1, m]| = A(m) \geq \alpha m.$$

(This is also true for $m = 0$, which may be the case for the last block.) Adding these estimates for the blocks above we obtain

$$\alpha\big((b_2 - b_1 - 1) + (b_3 - b_2 - 1) + \cdots + (b_k - b_{k-1} - 1) + (n - b_k) = \alpha(n - k).$$

Consequently

$$C(n) \geq k + \alpha(n - k) = \alpha n + (1 - \alpha)k \geq \alpha n + (1 - \alpha)\beta n = (\alpha + \beta - \alpha\beta)n.$$

□

Clearly this inequality also holds in the degenerate case when $\alpha = 0$, provided $0 \in A$.

PREXERCISE. Construct sets A, B satisfying $0 \in A$, $0 < \sigma(A), \sigma(B) < 1$ and

$$\sigma(A + B) = \sigma(A) + \sigma(B) - \sigma(A)\sigma(B).$$

PREXERCISE. Show that for every pair of sets satisfying the previous exercise the values of $\sigma(A)$ and $\sigma(B)$ are always rational.

We can write (2.1) in the symmetric form

$$1 - \sigma(A + B) \leq \big(1 - \sigma(A)\big)\big(1 - \sigma(B)\big).$$

An iterated application then gives

$$1 - \sigma(hA) \leq \big(1 - \sigma(A)\big)^h, \tag{2.2}$$

which will become small but not quite 0. We complement this inequality with the following result.

THEOREM 2.2. *Let A and B be sets of nonnegative integers with positive Schnirelmann densities $\sigma(A) = \alpha$ and $\sigma(B) = \beta$, respectively. If $\alpha + \beta \geq 1$ and $0 \in A \cup B$, then $A + B \supset \mathbb{N}$.*

PROOF. Assume $0 \in A$ and take a positive integer n. We want to prove $n \in A+B$. If $n \in B$, we are done, so assume now $n \notin B$. This implies

$$B(n-1) = B(n) \geq \beta n.$$

Consider the pairs $(i, n-i)$ with $1 \leq i \leq n-1$. In $A(n-1)$ cases we have $i \in A$, and in $B(n-1)$ cases we have $n-i \in B$. Since

$$A(n-1) + B(n-1) \geq \alpha(n-1) + \beta n > n-1,$$

at least once both happen. □

We can now prove Schnirelmann's theorem on bases.

PROOF OF THEOREM 1.4. Assume $\sigma(A) = \alpha > 0$. Take an integer h such that $(1-\alpha)^h < 1/2$. Then $\sigma(hA) > 1/2$ according to (2.2), and so $2hA \supset \mathbb{N}$ by the previous theorem. □

The above argument estimates the order of this basis by $(\log 4)/\alpha$. We will see that the optimal estimate is $1/\alpha$.

EXERCISE 75. Given an integer h, construct a set A such that $0 \in A$, $\sigma(A) = 1/h$ and the exact order of A is h.

3. Mann's theorem

In Schnirelmann's theorem the role of the sets A, B is asymmetric: one of them contains 0 and the other need not. We now show how this inequality can be improved under the symmetric condition $0 \in A \cap B$, which is also better suited for the repeated addition of the same set.

THEOREM 3.1 (Mann). *If $0 \in A \cap B$, then*

$$\sigma(A+B) \geq \min\big(1, \sigma(A) + \sigma(B)\big).$$

By iteration, if $0 \in A$, then $\sigma(kA) \geq \min\big(1, k\sigma(A)\big)$ and thus, next corollary follows.

COROLLARY 3.2. *If $0 \in A$ and $\sigma A = \alpha > 0$, then A is a basis of order $\leq 1/\alpha$.*

EXERCISE 76. Let α, β, γ be positive real numbers such that $\alpha + \beta \leq \gamma \leq 1$. Construct sets of positive integers such that $\sigma(A) = \alpha$, $\sigma(B) = \alpha$, $\sigma(A+B) = \gamma$. (Lepson [**29**].)

This theorem is similar to the Cauchy-Davenport inequality: superadditivity save an obstruction, which in our case consists in densities being bounded by 1. The proof will also be based on a transfusion method. However, while a transfusion preserves the sum of cardinalities it does typically change the sum of densities. It does not change the value of $A(n) + B(n)$ for any n, and this suggests the following approach.

DEFINITION 3.3. The *joint (Schnirelmann) density* of the sets $A_1, \ldots, A_k$ is defined by

$$\sigma(A_1, \ldots, A_k) = \inf \frac{A_1(n) + \cdots + A_k(n)}{n}.$$

Now Theorem 3.1 will follow from the version below.

THEOREM 3.4. *If $0 \in A \cap B$, then*

$$\sigma(A+B) \geq \min\bigl(1, \sigma(A,B)\bigr).$$

Formally Schnirelmann density is a limit, a thing related to infinity, but it gives information for every $A(n)$; it is perhaps not surprising that the above theorem will be proved in a finite setting.

THEOREM 3.5. *Let $0 \leq \gamma \leq 1$, let n be a positive integer and let A, B be sets such that $0 \in A \cap B$. Put $C = A + B$. If*

$$A(k) + B(k) \geq \gamma k \text{ for } 1 \leq k \leq n, \tag{3.1}$$

then

$$C(k) \geq \gamma k \text{ for } 1 \leq k \leq n. \tag{3.2}$$

EXERCISE 77. Deduce Theorem 3.5 from Theorem 3.4.

We present another slight (but useful) generalization.

THEOREM 3.6. *Let $0 \leq \gamma \leq 1$, $0 \leq \delta \leq 1 - \gamma$, let n be a positive integer and let A, B be sets such that $0 \in A \cap B$. Put $C = A + B$. If*

$$A(k) + B(k) \geq \gamma k - \delta \text{ for } 1 \leq k \leq n, \tag{3.3}$$

then

$$C(k) \geq \gamma k - \delta \text{ for } 1 \leq k \leq n. \tag{3.4}$$

The other interesting case is $\delta = 1 - \gamma$, which can be reformulated as follows.

THEOREM 3.7 (Van der Corput). *Let $0 \leq \gamma \leq 1$, let n be a positive integer and let A, B be sets such that $0 \in A \cap B$. Put $C = A + B$. If*

$$1 + A(k) + B(k) \geq \gamma(k+1) \text{ for } 1 \leq k \leq n, \tag{3.5}$$

then

$$1 + C(k) \geq \gamma(k+1) \text{ for } 1 \leq k \leq n. \tag{3.6}$$

PROOF. Suppose the above statement is false; then among the counterexamples there is one with the smallest value of n, and with n fixed, with the minimal value of $B(n)$. We consider this example now. We may assume that

$$A, B \subset [0, n],$$

since omitting the element outside this range does not change the assumptions or the conclusion.

If $n = 1$, then either $1 \in A \cup B$ and then $C(1) = 1 \geq \gamma - \delta$, or $1 \notin A \cup B$ and $\gamma = 0$, so $C(1) = 0 \geq -\delta$. So assume $n \geq 2$.

If $B = \{0\}$, then the statement is obviously true. Assume B has also positive elements.

We try to make a translation-transfusion in the following form: we try to replace A, B by

$$A' = A \cup (B + t),\ B' = B \cap (A - t)$$

with suitable t. Any such pair of sets satisfies $A' + B' \subset A + B$. Hence this will also be a counterexample, provided it satisfies conditions $0 \in A \cap B$ and (3.3). The first is equivalent to $t \in A$; we will return to the second.

This pair of sets will contradict the minimality assumption if $B' \neq B$, that is,

$$B' \not\subset A - t.$$

Such values of t do exist, for instance, the maximal element of A has this property. From such values of t we choose the minimal one. The minimality of t means that

$$a \in A, a < t \Longrightarrow B \subset A - a \Longrightarrow B + a \subset A. \tag{3.7}$$

A consequence of (3.7) (with $a = 0$) is

$$B \subset A.$$

Another one is

$$A(x) = C(x) \geq \gamma x - \delta \text{ for } x < t.$$

Indeed, (3.7) means that any sum of the form $b + a$, $b \in B, a \in A$ with $a < t$ is in A, and this includes all cases when $a + b < t$. Furthermore, restricting the inclusion (3.7) for a fixed $b \in B$ and $a \leq x < t$ we see that

$$b + \big(A \cap [0, x]\big) \subset A \cap [b, b + x].$$

Comparing the cardinalities we obtain

$$A(b + x) - A(b - 1) \geq A(x) + 1 = C(x) + 1 \geq \gamma x - \delta + 1 \tag{3.8}$$

for $x < t$.

Our aim is to show that

$$A'(k) + B'(k) \geq \gamma k - \delta \text{ for } 1 \leq k \leq n. \tag{3.9}$$

From the definition of A', B' we immediately see that

$$A'(k) + B'(k) \geq A(k) + B(k - t),$$

hence this holds if $B(k - t) = B(k)$, in particular, if $t = 0$. So we may assume that $t > 0$ and $B(k - t) < B(k)$. This means that there are elements of B in the interval $(k - t, k]$; let b' be the smallest of them. Write $k = b' + x$, $0 \leq x < t$.

We have

$$\begin{aligned} A'(k) + B'(k) &\geq A(k) + B(k - t) = A(k) + B(b' - 1) \\ &= \big(A(b' - 1) + B(b' - 1)\big) + \big(A(k) - A(b' - 1).\big) \end{aligned}$$

We estimate the first term by the induction hypothesis, the second one by (3.8):

$$A'(k) + B'(k) \geq \big(\gamma(b' - 1) + \delta\big) + \big(\gamma(k - b') + 1 - \delta.\big) = \gamma k + 1 - \gamma - 2\delta \geq \gamma k - \delta;$$

in the last step we need the assumption $\gamma + \delta \leq 1$. □

4. Schnirelmann's theorem revisited

In Schnirelmann's theorem 1.4 equality can hold for *certain* values of α and β (Exercise 2). Lepson [**29**] showed that in Mann's Theorem 3.1 equality can hold for any α and β (Exercise 76 above).

By writing

$$S(\alpha, \beta) = \inf\{\sigma(A + B) : \sigma(A) = \alpha,\ \sigma(B) = \beta,\ 0 \in A\}$$

and

$$M(\alpha, \beta) = \inf\{\sigma(A + B) : \sigma(A) = \alpha,\ \sigma(B) = \beta,\ 0 \in A \cap B\},$$

we can restate Schnirelmann's, Mann's and Lepson's results as

$$\alpha+\beta-\alpha\beta \le S(\alpha,\beta) \le M(\alpha,\beta) = \min(\alpha+\beta,1). \tag{4.1}$$

If $\alpha+\beta>1$, then we have $S(\alpha,\beta)=M(\alpha,\beta)=1$, (Schnirelmann's Theorem 2.2), thus in both inequalities of (4.1) equality can actually occur.

In this section we give a formula for $S(\alpha,\beta)$ and describe the cases of equality in inequalities (4.1). These results are from Hegedűs-Piroska-Ruzsa [**21**].

THEOREM 4.1. *For all α,β we have*

$$S(\alpha,\beta) = \inf_{n\ge 0} \frac{\lceil \alpha n\rceil + \lceil \beta(n+1)\rceil}{n+1}. \tag{4.2}$$

DEFINITION 4.2. Let α,β be positive real numbers satisfying $\alpha+\beta\le 1$. We call (α,β) a *Schnirelmann pair* if $S(\alpha,\beta)=\alpha+\beta-\alpha\beta$, and a *Mann pair* if $S(\alpha,\beta)=\alpha+\beta$.

THEOREM 4.3. *The numbers (α,β) form a Schnirelmann pair if and only if they can be expressed as*

$$\alpha=\frac{k}{n},\quad \beta=\frac{1}{n+1}$$

with certain integers $n\ge 2$ and $1\le k\le n-1$.

PROOF OF THEOREM 4.1. Denote the right side of (4.2) by γ. First we show that $S(\alpha,\beta)\ge\gamma$. Since $\sigma(B)>0$, we have $1\in B$. Write $B'=B-1$; thus $0\in B'$. We will apply the above lemma to the sets A,B'; the requirement that both contain 0 is hence fulfilled.

Next we show that the sets A,B' satisfy (3.5). Indeed, by the definition of the Schnirelmann density we have $A(k)\ge\alpha k$, and since it must be an integer, we have

$$A(k)\ge\lceil\alpha k\rceil.$$

We have

$$B'(k)=|B'\cap[1,k]|=|B\cap[2,k+1]|=B(k+1)-1\ge\beta(k+1)-1,$$

and again this is an integer, thus

$$B'(k)\ge\lceil\beta(k+1)\rceil-1.$$

On adding these inequalities we find

$$1+A(k)+B'(k)\ge\lceil\alpha k\rceil+\lceil\beta(k+1)\rceil\ge\gamma(k+1)$$

by the definition of γ.

An application of Theorem 3.7 to the sets A,B' yields that their sum $C'=A+B'$ satisfies

$$1+C'(n)\ge\gamma(n+1)$$

for all n.

Since $C=A+B$ is connected to C' via $C=C'+1$, we conclude that

$$C(n)=|C\cap[1,n]|=|C'\cap[0,n-1]|=1+C'(n-1)\ge\gamma n$$

for all n, which is equivalent to saying $\sigma(C)\ge\gamma$.

To show that $S(\alpha,\beta)\le\gamma$, suppose first that the infimum in the definition (1.4) is a minimum, and let n be any integer satisfying

$$\gamma=\frac{\lceil\alpha n\rceil+\lceil\beta(n+1)\rceil}{n+1}.$$

Consider the sets

$$A_0 = \{0, 1, \dots, \lceil \alpha n \rceil\} \cup \{n+1, n+2, \dots\}$$

and

$$B_0 = \{1, \dots, \lceil \beta(n+1) \rceil\} \cup \{n+2, n+3, \dots\}.$$

These sets satisfy

$$\sigma(A_0) = \frac{\lceil \alpha n \rceil}{n} \geq \alpha$$

and

$$\sigma(B_0) = \frac{\lceil \beta(n+1) \rceil}{n+1} \geq \beta.$$

We can select subsets $A \subset A_0$ and $B \subset B_0$ such that $\sigma(A) = \alpha$, $\sigma(B) = \beta$ and $0 \in A$. These sets satisfy

$$A + B \subset A_0 + B_0 = \{1, 2, \dots, \lceil \alpha n \rceil + \lceil \beta(n+1) \rceil\} \cup \{n+2, \dots\},$$

consequently (by evaluating the counting function at $n+1$) we find that

$$\sigma(A+B) \leq \sigma(A_0 + B_0) \leq \frac{\lceil \alpha n \rceil + \lceil \beta(n+1) \rceil}{n+1} = \gamma$$

as wanted.

Suppose next that the infimum is not attained. In this case we have

$$s = \inf_{n \geq 0} \frac{\lceil \alpha n \rceil + \lceil \beta(n+1) \rceil}{n+1} = \lim_{n \to \infty} \frac{\lceil \alpha n \rceil + \lceil \beta(n+1) \rceil}{n+1} = \alpha + \beta,$$

hence the example of equality in Mann's theorem serves also as an example for $S(\alpha, \beta) \leq \gamma$. □

Prof of Theorem 4.3. By Theorem 4.1, α and β form a Schnirelmann pair if and only if

$$\inf_{n \geq 0} \frac{\lceil \alpha n \rceil + \lceil \beta(n+1) \rceil}{n+1} = \alpha + \beta - \alpha\beta. \tag{4.3}$$

Since the limit of the left side of (4.3) is $\alpha + \beta$, in this case there must be an n such that

$$\frac{\lceil \alpha n \rceil + \lceil \beta(n+1) \rceil}{n+1} = \alpha + \beta - \alpha\beta.$$

Observe that the value of the left side for $n = 0$ is 1, so we must have $n \geq 1$. Write

$$\lceil \alpha n \rceil = k, \quad \lceil \beta(n+1) \rceil = l.$$

We have $\alpha n \leq k$ and $\beta(n+1) \leq l$, hence $k \neq 0$, $l \neq 0$ and

$$\alpha \leq k/n, \quad \beta \leq l/(n+1). \tag{4.4}$$

By the monotonicity of the function $\alpha + \beta - \alpha\beta$ in both variables (in our domain), we have

$$\alpha + \beta - \alpha\beta \leq \frac{k}{n} + \frac{l}{n+1} - \frac{k}{n}\frac{l}{n+1} = \frac{k+l}{n+1} - \frac{k(l-1)}{n(n+1)}.$$

Since $l \geq 1$, the last expression is always $\leq (k+l)/(n+1)$, and equality can hold only if $l = 1$ and both inequalities in (4.4) hold with equality. This means that $\alpha = k/n$ and $\beta = l/(n+1) = 1/(n+1)$ as claimed. □

We mention without proof some results on Mann pairs.

THEOREM 4.4. *If α and β form a Mann pair, then they are either both rational or both irrational. A pair of rational numbers, say $\alpha = p/q$, $\beta = r/s$ is a Mann pair if and only if they satisfy*

$$\{\alpha(1-n)\} + \{-\beta n\} \geq \alpha \tag{4.5}$$

for every integer $1 \leq n \leq \operatorname{lcm}[q,s]$. A pair of irrational numbers is a Mann pair if and only if there are integers k, l, m such that

$$\alpha k + \beta l = m, \quad 0 < k < 1/\alpha,\ 0 \leq k - l < 1/\alpha. \tag{4.6}$$

The description of rational Mann pairs is less satisfactory than that of irrational ones, though it provides a finite algorithm for each pair of rational numbers. The following can be observed.

STATEMENT 4.5. *Let α, β be rational numbers, and write $\alpha/\beta = a/b$ with $(a,b) = 1$. If there are integers satisfying (4.6) , then (α, β) is a Mann pair. In particular, if $\alpha \leq 1/(a+b)$, then it is a Mann pair.*

The difficulty is that the set P will be a lattice in the rational case, and there seems no easy way to decide when a lattice intersects a triangle.

We note that condition (4.6) is not necessary in the rational case. This is seen by the examples $\alpha = 4/11$, $\beta = 5/11$ or $\alpha = 8/65$, $\beta = 2/13$.

5. Kneser's theorem, density form

In the previous sections we discussed addition theorems based on Schnirelmann density. An analogous result was found for the more natural concept of asymptotic density by Kneser [**26**]. This is more complicated than the privous ones and its proof is difficult. We state it without proof; a proof can be found in Halberstam and Roth's monograph [**20**].

THEOREM 5.1 (Kneser). *Let A and B be sets of positive integers. Either*

$$\underline{d}(A+B) \geq \underline{d}(A) + \underline{d}(B),$$

or there exists positive integers q, k, l such that $q \geq k + l - 1$ and
(a) A is contained in k residue classes modulo q,
(b) B is contained in l residue classes modulo q,
(c) $A + B$ is equal to $k + l - 1$ residue classes modulo q except a finite set.

A density cannot exceed 1; in the above formulation the case when $\underline{d}(A) + \underline{d}(B) > 1$ is included as the extremal case $q = k = l = 1$.

A typical example of the second case case is $A = \{1, \dots, k \mod q\}$, the first k residue classes modulo q and $B = \{1, \dots, l \mod q\}$, the first l classes modulo q.

EXERCISE 78. Suppose $\underline{d}(A) + \underline{d}(B) = 1$. Show that $A + B$ has an asymptotic density, and find its possible values.

6. Adding a basis: Erdős' theorem

The previous results gave estimates for the density of a sumset using the density of summands. Sometimes a density increment occurs also when we add a set of density 0. The first example of this phenomenon was given by Hinchin in 1933. He proved that for the set Q of nonnegative squares we have $\sigma(A+Q) > \sigma(A)$ whenever $0 < \sigma(A) < 1$.

A few years later Erdős proved that every basis has this property. In this section we give an account of this result.

THEOREM 6.1. *Let $B \subset \mathbb{Z}$ be a basis of order k and let $A \subset \mathbb{Z}$. Then*

$$\sigma(A+B) \geq \sigma(A) + \frac{\sigma(A)\,(1-\sigma(A))}{2k}.$$

PROOF. Write $\alpha = \sigma(A)$ and let $C = A + B$. We are going to estimate $C(n)$.

We will try to find a $b \in B$ with a large proportion of $(A+b)$ outside A; this will make $A \cup A + b$ large. For that purpose define

$$f(t) = |((A+t) \setminus A) \cap [1,n]|\,.$$

This function has a subadditivity property. Indeed, we have

$$\begin{aligned}(A+x+y) \setminus A &\subset \big((A+x+y) \setminus (A+x)\big) \cup \big((A+x) \setminus A\big) \\ &= \big((A+y) \setminus A\big) + x \cup \big((A+x) \setminus A\big),\end{aligned}$$

and by comparing the cardinalities we easily find

$$f(x+y) \leq f(x) + f(y).$$

Observe that we have $f(0) = 0$. We are going to calculate the average of f. We have

$$\begin{aligned}\sum_{t=1}^{n-1} f(t) &= |\{(a,t) : 1 \leq a < a+t \leq n, a \in A, a+t \notin A\}| \\ &= |\{(a,x) : 1 \leq a < x \leq n, a \in A, x \notin A\}|\end{aligned}$$

by introducing $x = a + t$.

Since

$$\{(a,x), 1 \leq a < x \leq n, a \in A, x \notin A\} \cup \{(a,x), 1 \leq a < x \leq n, a \in A, x \in A\} =$$

$$= \{(a,x), 1 \leq a < x \leq n, a \in A\},$$

and the cardinality of this last set can be expressed by counting $\{(a,x), 1 \leq a < x \leq n, a \in A\}$ over x as $\sum A(x-1)$, we find

$$\sum_{t=1}^{n-1} f(t) = \sum_{x=1}^{n} A(x-1) - |\{(a,x), a \in A, x \in A, 1 \leq a < x \leq n\}|\,.$$

Using the definition of the Schnirelmann's density we can conclude that $A(x-1) \geq \alpha(x-1)$. As the second part is equal to $A(n)(A(n)-1)/2$, we can bound $\sum f(t)$ by

$$\sum_{t=1}^{n-1} f(t) \geq \alpha\frac{n(n-1)}{2} - \frac{A(n)(A(n)-1)}{2}.$$

This inequality implies that there exist a t_0 for which this $f(t_0)$ is large:

$$f(t_0) \geq \frac{1}{n-1}\left(\frac{\alpha n(n-1)}{2} - \frac{A(n)(A(n)-1)}{2}\right).$$

Since B is a basis of order k, we can write $t_0 = b_1 + \ldots + b_k$, for some $b_j \in B$. From the subadditivity property we conclude $f(t_0) \leq \sum f(b_i)$, consequently there is a $b = b_i$ for which

$$f(b) \geq \frac{1}{k}\frac{1}{n-1}\left(\frac{\alpha n(n-1)}{2} - \frac{A(n)(A(n)-1)}{2}\right).$$

In particular, as $C(n) \geq A(n) + f(b)$ for any single b, we get

$$C(n) \geq A(n) + \frac{1}{k}\frac{1}{n-1}\left(\frac{\alpha n(n-1)}{2} - \frac{A(n)\,(A(n)-1)}{2}\right).$$

Since the right hand side, as a function of $A(n)$, is increasing up to $A(n) \leq k(n-1)+1/2$, we get a lower estimate by replacing each occurrence of $A(n)$ by its lower bound αn, and we obtain

$$C(n) \geq \alpha n + \frac{1}{k(n-1)}\left(\frac{\alpha n(n-1)}{2} - \frac{\alpha n(\alpha n-1)}{2}\right) \geq \alpha n + \frac{\alpha\,(1-\alpha)\,n}{2k}.$$

As this fact is true for all n, the estimate for the Schnirelmann's density follows. □

With a minimal modification of the proof a similar result can be obtained for asymptotic lower density.

THEOREM 6.2. *Let $B \subset \mathbb{Z}$ be an asymptotic basis of order k and let $A \subset \mathbb{Z}$. Then*

$$\underline{d}(A+B) \geq \underline{d}(A) + \frac{\underline{d}(A)\,(1-\underline{d}(A))}{2k}.$$

7. Adding a basis: Plünnecke's theorem, density form

If we add a basis of order k to a set of density $\alpha > 0$, Erdős' theorem in the previous section estimates the density of the sumset essentially by $\alpha(1+1/(2k))$ for small values of α. Plünnecke [**39**] gave a much stronger estimate, one which goes to infinity after division by α as $\alpha \to \infty$.

THEOREM 7.1. *If $A, B \subset \mathbb{N}_0$, $0 \in B$, then*

$$\sigma(A+B) \geq \sigma(A)^{1-\frac{1}{k}}\sigma(kB)^{\frac{1}{k}}. \tag{7.1}$$

In particular, if $kB = \mathbb{N}_0$, then $\sigma(A+B) \geq \sigma(A)^{1-\frac{1}{k}}$.

EXERCISE 79. Prove that

$$\alpha^{1-1/k} > \alpha + \frac{\alpha(1-\alpha)}{k}$$

for all $\alpha \in (0,1)$, hence Plünnecke's inequality is stronger than that of Erdős for all possible values of $\sigma(A)$.

Plünnecke's theorem gives us the correct order of magnitude.

EXERCISE 80. Construct a basis B of order k, such that, for all α, exist a set A, with $\sigma(A) \geq \alpha$, such that $\sigma(A+B) < C\alpha^{1-\frac{1}{k}}$, for some $C \geq 0$.

PROOF. Before starting the proof we remark that a result which is weaker by a constant factor can be easily deduced from the finite Plünnecke inequality given in Chapter 1. Indeed, the case $j = 1$ of Corollary 2.4, or the case $l = 0$ of Theorem 1.1 of

Chapter 1 gives us the following. If A, B are finite sets and $|A| = m$, $|A+B| = \alpha m$, then $|kB| \leq \alpha^k m$. By substituting $\alpha = |A+B| / |A|$ and rearranging this can be written as

$$|A+B| \geq |A|^{1-1/k} |B|^{1/k}. \tag{7.2}$$

This is analogous to (7.1), just we have cardinality here and density there.

Let now A, B be infinite sets such that A and kB have positive Schnirelmann density. Let $C = A + B$; we want to estimate $C(n)$. To this end we apply (7.2) for the sets

$$A' = A \cap [1, \lceil n/2 \rceil], \ B' = B \cap [0, [n/2]].$$

We have $|A'| \geq \sigma(A)n/2$. Since the set kB' contains every element of kB up to $n/2$ (and contains 0), we have $|kB'| \geq \sigma(kB)n/2$. Now an application of (7.2) gives

$$C(n) \geq |A' + B'| \geq \sigma(A)^{1-\frac{1}{k}} \sigma(kB)^{\frac{1}{k}} n/2.$$

Since this holds for every n, we have a lower estimate for $\sigma(A+B)$, which is half of the one claimed in (7.1).

To remove this factor we apply an induction argument like for Mann's theorem, and use Plünnecke's method for a trimmed additive graph.

We write $\alpha = \sigma(A)$, $\beta = \sigma(kB)$, $C = A + B$ and $\gamma = \alpha^{1-\frac{1}{k}} \beta^{\frac{1}{k}}$. We want to show that $C(n) \geq \gamma n$ for all n.

We reformulate this in the following finite form.

Let $n \in \mathbb{N}$. If $A(m) \geq \alpha m$ for all $m \leq n$, $0 \in B$ and $B_k(m) = |kB \cap [1, m]| \geq \beta m$ for all $m \in \mathbb{N}$, then $C(m) \geq \gamma m$ for all $m \leq n$.

We will proceed by induction on n. Since the case $n = 1$ is clear, suppose that $n > 1$ and that the claim above has been proved for all $n' < n$. Let

$$A_1 = A \cap [1, n']$$

$$A_2 = A \cap [n'+1, n] - n' \subset [1, n - n'].$$

Observe that $A_1(m) \geq \alpha m$ is satisfied for all $m \leq n'$. If, for some $n' < n$, it also happens that $A_2(m) \geq \alpha m$ for all $m \leq n - n'$, then we apply the induction hypothesis for the set pairs A_1, B and A_2, B. Since

$$C \supset (A_1 + B) \cup (A_2 + n' + B),$$

we get the desired conclusion for C.

Now consider the case when this does not happen. This means that for all $n' < n$ there is some $m \leq n - n'$ such that

$$|A \cap [n'+1, n'+m]| < \alpha m.$$

We claim that the above inequality is satisfied for $m = n - n'$, i.e.,

$$|A \cap [n'+1, n]| < \alpha(n - n'). \tag{7.3}$$

Put $n_1 = n'$. We find m_1 such that

$$|A \cap [n_1 + 1, n_1 + m_1]| < \alpha m_1.$$

If $m_1 = n - n'$ we are done. If not, we take $n_2 = n' + m_1$ and obtain m_2 such that

$$|A \cap [n_2 + 1, n_2 + m_2]| < \alpha m_2.$$

We iterate this process and when it stops, we add all the inequalities to get (7.3).

Now we build a restricted addition graph (like described in Section 2 of Chapter 1) as follows. It consists of the layers $V_0 = A \cap [1, n], V_1 = (A + B) \cap [1, n], \ldots, V_h =$

$(A+hB)\cap[1,n]$, with edges going from each $x\in V_{i-1}$ to $x+b\in V_i$ as usual. One can easily to check that it is commutative, which allows us to apply Plünnecke's inequality (Theorem 2.1 of Chapter 1). We use μ_j to denote the magnification ratios of this graph. This inequality tells us $\mu_k\le\mu_1^k$. For μ_1 we use the obvious estimate $\mu_1\le C(n)/A(n)$ to deduce

$$\mu_k\le\left(\frac{C(n)}{A(n)}\right)^k. \tag{7.4}$$

Now we find a lower bound for μ_k. Let $X\subset A$ be such that $|\mathrm{im}(X,V_k)|=\mu_k|X|$, and let $n'+1$ be its first element, so that $X\subset[n'+1,n]$. From (7.3) we infer

$$|X|\le\alpha(n-n')$$

if $n'>0$, while for $n'=0$ we have $|X|\le A(n)$.

We also know,

$$|(X+kB)\cap[1,n]|\ge|(n'+1+kB)\cap[n'+1,n]|\ge\beta(n-n').$$

Combining the two inequalities we get

$$\mu_k\ge\min\left\{\frac{\beta}{\alpha},\frac{\beta n}{A(n)}\right\}.$$

Now (7.4) completes the proof. □

We can easily deduce a similar asymmetric combining lower asymptotic density and Schnirelmann density.

THEOREM 7.2. *Let $A,B\subset\mathbb{Z}$ and let k be a positive integer. We have*

$$\underline{d}(A+B)\ge\underline{d}(A)^{1-\frac{1}{k}}\sigma(kB)^{\frac{1}{k}}$$

This can be deduced from the previous theorem using the connection between lower and Schnirelmann density as expressed in Exercise 73.

Let $\epsilon>0$. To prove this result we find a t such that $\sigma(A-t)\ge\underline{(}A-\epsilon)$. This implies that

$$\underline{d}(A+B)\ge\sigma(A+B-t)\ge(\underline{d}(A)-\epsilon)^{1-\frac{1}{k}}\,\sigma(kB)^{\frac{1}{k}}.$$

The result follows by letting ϵ go to 0.

The theorem is also true with the lower density everywhere but the proof is more involved.

8. Adding the set of squares or primes

Let Q be the set of squares: $Q=\{n^2,n\in\mathbb{N}_0\}$. They form a basis of order 4, that is, $4Q=\mathbb{N}_0$. We also know that $3Q$ contains almost number except those of the form $4^a(8b-1)$, for some a and b. This easily implies that the Schnirelmann density of the set of three-fold sums of squares is positive: $\sigma(3Q)>0$.

EXERCISE 81. Calculate $d(3Q)$ and $\sigma(3Q)$.

If we have $A\subset\mathbb{Z}$ with $\sigma(A)=\alpha$, then using Plünnecke's density theorems from the previous section we see that $\sigma(A+Q)\ge\alpha^{\frac{3}{4}}$, and for small α we can improve this to $\sigma(A+Q)>c\alpha^{\frac{2}{3}}$. This is still not the best possible; the real exponent is $\frac{1}{2}$.

THEOREM 8.1 (Plünnecke [**37**]). *Let A a subset of the integers with $\sigma(A) = \alpha$, and Q the set of squares. We have*

$$\sigma(A+Q) \geq c\alpha^{\frac{1}{2}}$$

for some absolute constant $c > 0$.

To see that this exponent is sharp, set $A = \{1, q+1, q+2, \ldots\}$, with a large q. This set has Schnirelmann density $\sigma(A) = \frac{1}{q}$. Since up to q the only elements of $A+Q$ are the integers of form k^2+1, we easily find that

$$\sigma(A+Q) = \frac{1+[\sqrt{q}]}{q} \sim \frac{1}{\sqrt{q}} = \sqrt{\sigma(A)}.$$

For asymptotic density the increase is larger, a similar result holds with arbitrarily small exponent.

THEOREM 8.2 ([**45**]). *For every $\epsilon > 0$ there exists a constant c_ϵ depending on ϵ such that if $\underline{d}(A) = \alpha$ then*

$$\underline{d}(A+Q) \geq c_\epsilon \alpha^\epsilon.$$

A good configuration that gives approximately the correct order of magnitude is a residue class modulo some q. Put $A = q\mathbb{Z}$, so that $\underline{d}(A) = \frac{1}{q}$. The sumset $A+Q$ contains the quadratic residues modulo q. If the prime factorization of q is $q = p_1 \ldots p_k$, then the number of quadratic residues is

$$\left(\frac{p_1+1}{2}\right) \cdots \left(\frac{p_k+1}{2}\right) \approx \frac{q}{2^k}.$$

To minimize it we take the product of the first primes. By the prime number theorem this will be essentially the primes $p_i < \log q$, and their number is $\sim (\log q)/(\log\log q)$. So the density of $A+Q$ is approximately

$$\underline{d}(A+Q) \approx \frac{1}{q} \frac{q}{2^{\frac{\log q}{\log\log q}}} = q^{-\frac{\log_q 2 \log q}{\log\log q}}.$$

The exponent goes that can go to zero as q grows.

Let P be the prime numbers, and let $P' = P - 2$. It is known that there exists a k such that $kP' = \mathbb{N}_0$ for some k ($k = 7$?). It is also known that $d(2P) = 1/2$, $d(3P) = 1$ and that every large number can be written as a sum of four primes. These give estimates for the density of $A+P$ by the Plünnecke inequalities; similarly to the case of squares, this is far from the reality.

THEOREM 8.3. [**45, 46**] *Let A be a subset of integers. There is a positive constant c with the following properties (valid for q sufficiently large).*
(a) If $\sigma(A) = \frac{1}{q}$ then $\sigma(A+P') \geq \frac{c}{\log q}$
(b) If $\underline{d}(A) = \frac{1}{q}$ then $\underline{d}(A+P) \geq \frac{c}{\log\log q}$

The examples to show that this is the correct order of magnitude are similar to the case of squares. For Schnirelmann density use the same set $A = \{1, q+1, q+2, \ldots\}$. This set has Schnirelmann density $\sigma(A) = \frac{1}{q}$. Since up to q the only elements of $A+P'$ are neigbours of primes, we find that

$$\sigma(A+Q) = \frac{\pi(q+1)}{q} \sim \frac{1}{\log q}.$$

For asymptotic density again we use the multiples of q. The numbers in $A+P$ are sets coprime to q with finitely many exceptions. Then

$$d(A+P)=\frac{\varphi(q)}{q}=\left(1-\frac{1}{p_1}\right)\dots\left(1-\frac{1}{p_k}\right)\sim\frac{c}{\log\log q}$$

if again we take the product of the first primes.

9. Essential components

We say that B is an *essential component* if B is such that $\sigma(A+B)>\sigma(A)$ for every $0<\sigma(A)<1$.

In the previous sections we met examples of essential components: first sets of positive density (theorems of Schnirelmann and Mann), then bases (theorems of Erdős and Plünnecke). The first example of an essential component that is not a basis was given by Linnik; this set was too thin to be a basis. Clearly if B is a basis of order k, it must satisfy $B(x)>x^{1/k}$, hence a set such that $B(x)=O(x^{\varepsilon})$ for every positive ε cannot be a basis.

The following theorem tells exactly how thin an essential component can be.

THEOREM 9.1. [**44**]
(a) for every $\epsilon>0$ exist an essential component with $B(n)<c(\log n)^{1+\epsilon}$.
(b) There is no essential component B with $B(n)<c(\log n)^{1+o(1)}$.

CHAPTER 5

Measure and topology

1. Introduction

In this chapter we mention some loosely connected things. The common feature is that we now leave the safe familiar world of finite sets. Our excursions are in two different directions.

The first is to measures. Measure is a close analog of cardinality; the same questions we asked for finite sets can be formulated for measures of sets of reals, or in $\mathbb{R}^d$, or in a more general setting. We already mentioned a classical example, the Brunn-Minkowski inequality: for measurable sets in $\mathbb{R}^d$ we have

$$\mu(A+B)^{1/d} \geq \mu(A)^{1/d} + \mu(B)^{1/d}. \tag{1.1}$$

This illustrates some basic differences. While measure is a more sophisticated concept than cardinality of a finite sets, the results are often simpler and sometimes also easier to prove.

The second excursion is to topology. We think of the integers as a discrete set; however, other topologies on them do exist, and some have a relevance to our subject. We will mainly discuss the connection of the Bohr topology to additive properties.

2. Raikov's theorem and generalizations

A natural analog of adding integers modulo q is the addition of reals modulo 1. The analog of the Cauchy-Davenport inequality is the following theorem, due to Raikov [**40**] from 1939.

In the sequel we consider subsets of $[0,1)$, addition is meant modulo 1. Problems of mesurability are not in the focus of our interest, so assume that every set mentioned is compact or open. Lebesgue measure is denoted by μ.

THEOREM 2.1. *For $A, B \subset [0,1)$ we have*

$$\mu(A+B) \geq \min\bigl(1, \mu(A)+\mu(B)\bigr). \tag{2.1}$$

PREXERCISE. Deduce Raikov's theorem from the Cauchy-Davenport inequality.

a) Approximate a general set by a union of intervals.

b) Taking a prime p and compare $\mu(A)$ to the cardinalites of the sets

$$A(x,p) = \{j \in \mathbb{Z}_p : x + j/p \in A\}. \tag{2.2}$$

(Naturally j/p and $x+j/p$ are interpreted modulo 1.)

Of the two approaches suggested above, a) is more natural but also has more cumbersome details. We describe method b).

PROOF. Write $A+B=S$. Take a prime p and construct sets of residues from the sets A, B, S as described in (2.2). For every x, y we have

$$S(x,p) \supset A(x-y,p) + B(y,p).$$

An application of the Cauchy-Davenport inequality to these sets now gives

$$|S(x,p)| \geq \min\big(|A(x-y,p)| + |B(y,p)| - 1, p\big).$$

We can reformulate this as follows. Either $S(x,p) = \mathbb{Z}_p$, or

$$|S(x,p)| \geq |A(x-y,p)| + |B(y,p)| - 1 \tag{2.3}$$

for each y.

The average of such a cardinality is connected to the measure of the set in an immediate way:

$$\int_0^1 |A(x,p)dx = p\mu(A).$$

(Why?) By integrating both sides of (2.3) with respect to y we obtain

$$|S(x,p)| \geq p(\mu(A) + \mu(B)) - 1.$$

This holds unless $S(x,p) = \mathbb{Z}_p$; hence the following inequality always holds:

$$|S(x,p)| \geq \min\big(p(\mu(A) + \mu(B)) - 1, p\big).$$

Now integrating this inequality with respect to x and dividing by p we get

$$\mu(A+B) \geq \min\big(1, \mu(A) + \mu(B)\big) - 1/p.$$

As this holds for every prime p, (2.1) follows. □

Raikov's theorem was generalized by Macbeath [**31**] for the n-dimensional torus, by Shields [**58**] for connected commutative compact second countable groups, by Kneser [**27**] for commutative locally compact groups, and by Kemperman [**23**] for noncompact groups. We state below his result in less than complete generality to avoid discussing certain aspects of noncommutative groups and measurability.

Let G be a locally compact topological group. If G is compact, or commutative, and in some other cases too, it has an invariant measure μ, called *Haar measure*. Invariance means that we have

$$\mu(A + x) = \mu(x + A) = \mu(A)$$

for every measurable set and every $x \in G$. Without any condition we can only claim that there is a right-invariant μ_r satisfying $\mu_r(A + x) = \mu_r(A)$ and a left-invariant μ_l satisfying $\mu_l(x+A) = \mu_l(A)$. If these coincide, that is, an invariant Haar measure exists, the group is called *unimodular*. We shall state the unimodular case. Also we restrict our attention to measurable sets.

THEOREM 2.2. *Let G be a compact, connected group, $A, B \subset G$ measurable sets such that $A + B$ is also measurable. We have*

$$\mu(A+B) \geq \min\big(\mu(A) + \mu(B), \mu(G)\big).$$

THEOREM 2.3. *Let G be a locally compact, noncompact group which does not have any proper compact-open subgroup. Let $A, B \subset G$ be measurable sets such that $A + B$ is also measurable. We have*

$$\mu(A+B) \geq \mu(A) + \mu(B).$$

3. The impact function

Let G be a group with a Haar measure μ, and $B \subset G$ a measurable set. We define the *impact function* of B analogously to the finite situation by the f text

$$\xi_A(x) = \inf\{\mu(A+B) : B \subset G, \mu(B) = x\}.$$

EXERCISE 82. Let G be the interval $[0,1)$ with addition modulo 1. Prove the concavity of the impact function: for $0 < y < x < 1$ and $x + y \leq 1$ we have

$$\xi(x-y) + \xi(x+y) \leq 2\xi(x).$$

EXERCISE 83. Use the previous exercise to give another proof of Raikov's theorem.

This method also can be extended to every locally compact group.

THEOREM 3.1. *If G does not have any proper compact-open subgroup, then ξ is a continuous concave function on its whole domain.*

See [**51**]; there this result is also applied to deduce Kemperman's theorems from the previous section.

Another curious property of the impact function is its symmetry. To avoid an exception we redefine the impact function at 0 by continuity:

$$\xi(0) = \lim_{x \to 0^+} \xi(x).$$

THEOREM 3.2. *Assume that G is compact, commutative and connected. The smallest value of x for which $\xi(x) = \mu(G)$ is $x = \mu(G) - \xi(0)$. The graph of $\xi(x)$ on the interval $[0, \mu(G) - \xi(0)]$ is symmetric to the line $x + y = \mu(G)$.*

4. Meditation on convexity and dimension

Let A, B be Borel sets in $\mathbb{R}^d$. The Brunn-Minkowski inequality (1.1) estimates $\mu(A+B)$ in a natural way, with equality if A and B are homothetic convex sets.

This can be exptessed in terms of the impact function as

$$\xi_B(a) \geq \left(a^{1/d} + \mu(B)^{1/d}\right)^d,$$

and this is the best possible estimate in terms of $\mu(B)$ only.

To measure the degree of nonconvexity one can try to use the measure of the convex hull beside the measure of the set. This is analogous to the hull volume, and it is sufficient to describe the asymptotic behaviour of ξ.

THEOREM 4.1 ([**57**], Theorem 1.). *For every bounded Borel set $B \subset \mathbb{R}^d$ of positive measure we have*

$$\lim_{a \to \infty} \xi_B(a)^{1/d} - a^{1/d} = \mu(\operatorname{conv} B)^{1/d}.$$

This is the continuous analogue of Theorem 9.3 of Chapter 3, and there is an analogue to the effective version Theorem 9.5 as well.

Note that by considering sets homothetic to $\operatorname{conv} B$ we immediately obtain

$$\xi_B(a)^{1/d} \leq a^{1/d} + \mu(\operatorname{conv} B)^{1/d},$$

thus we need only to give a lower estimate. This is as follows.

THEOREM 4.2 ([**57**], Theorem 2.). *Let $\mu(B) = b$, $\mu(\operatorname{conv} B) = v$. We have*

$$\xi_B(a)^{1/d} \geq a^{1/d} + v^{1/d}\left(1 - c(v/b)^{1/2}(v/a)^{1/(2d)}\right)$$

$$\xi_B(a) \geq a + dv^{1/d}a^{1-1/d}\left(1 - c(v/b)^{1/2}(v/a)^{1/(2d)}\right)$$

with a suitable positive constant c depending on d.

If $v > b$, we get a nontrivial improvement over the Brunn-Minkowski inequality for $a > a_0(b, v)$. It would be desirable to find an improvement also for small values of a, or, even more, to find the best estimate in terms of $\mu(B)$ and $\mu(\operatorname{conv} B)$.

The exact bound and the structure of the extremal set may be complicated. This is already so in the case $d = 1$, which was solved in [**49**]. Observe that in one dimension $\mu(\operatorname{conv} B)$ is is the diameter of B.

THEOREM 4.3 ([**49**], Theorem 2). *Let $B \subset \mathbb{R}$, and write $\mu(B) = b$, $\mu(\operatorname{conv} B) = v$. If*

$$a \geq \frac{v(v-b)}{2b} + \frac{b\{v/b\}(1 - \{v/b\})}{2}, \tag{4.1}$$

then $\xi_B(a) = a + v$. If (4.1) does not hold, then let k be the unique positive integer satisfying

$$\frac{k(k-1)}{2} \leq \frac{a}{b} < \frac{k(k+1)}{2}$$

and define δ by

$$\frac{a}{b} = \frac{k(k-1)}{2} + \delta k.$$

We have

$$\xi_B(a) \geq a + (k + \delta)b,$$

and equality holds if $B = [0, b] \cup \{v\}$.

A set A such that $\xi_B(a) = \mu(A + B)$ for the above set B is given by

$$A = [0, (k - 1 + \delta)b] \cup [v, v + (k - 2 + \delta)b] \cup \cdots \cup [(k-1)v, (k-1)v + \delta b].$$

A less exact, but simple and still quite good lower bound sounds as follows.

COROLLARY 4.4 ([**49**], Theorem 1). *Let $B \subset \mathbb{R}$, and write $\mu(B) = b$, $\mu(\operatorname{conv} B) = v$. We have*

$$\xi_B(a) \geq \min\bigl(a + v, (\sqrt{a} + \sqrt{b/2})^2\bigr).$$

A comparison with the 2-dimensional Brunn-Minkowski inequality gives the following iterpretation: initially a long one-dimensional set B tries to behave as if it were a two-dimensional set of area $b/2$.

It can be observed that Corollary 4.4 is weaker than the obvious inequality

$$\mu(A + B) \geq \mu(A) + \mu(B) \tag{4.2}$$

for small a. For small values of a Theorem 4.3 yields the following improvement of (4.2).

COROLLARY 4.5 ([**49**], Corollary 3.1). *If $a \leq b$, then we have*

$$\mu(A + B) \geq \min(2a + b, a + v).$$

If $b < a \leq 3b$, then we have

$$\mu(A + B) \geq \min\Bigl(\frac{3}{2}(a + b), a + v\Bigr).$$

PROBLEM 4.6. How large must $\mu(A+B)$ be if $\mu(A)$, $\mu(B)$, $\mu(\operatorname{conv} A)$ and $\mu(\operatorname{conv} B)$ are given?

What are the minima of $\mu(A+A)$ and $\mu(A-A)$ for fixed $\mu(A)$ and $\mu(\operatorname{conv} A)$?

The results above show that for $d=1$ (like in the discrete case, but for less obvious reasons) the limit relation becomes an equality for $a > a_0$. Again, this is no longer the case for $d=2$.

An example of a set $B \subset \mathbb{R}^2$ such that

$$\xi_B(a)^{1/2} < a^{1/2} + v^{1/2}$$

will hold for certain arbitrarily large values of a is as follows.

Let $0 < c < 1$ and let B consist of the square $[0,c] \times [0,c]$ and the points $(0,1), (1,0)$ and $(1,1)$. Hence $b = c^2$ and $v = 1$.

For an integer $n \geq 1$ put

$$A_n = [0,n] \times [0,n] \cup \bigcup_{j=0}^{n} [j, j+c] \times [n, n+c] \cup \bigcup_{j=0}^{n-1} [n, n+c] \times [j, j+c].$$

Thus A_n consists of a square of side n and $2n+1$ small squares of side c, hence

$$\mu(A_n) = n^2 + (2n+1)b.$$

We can easily see that $A_n + B = A_{n+1}$. Hence by considering the set $A = A_n$ we see that for a number a of the form $a = n^2 + (2n+1)b$ we have

$$\xi_B(a) \leq \mu(A_{n+1}) = (n+1)^2 + (2n+3)b < \left(\sqrt{a} + 1\right)^2.$$

A more detailed calculation leads to

$$\xi_B(a)^{1/2} \leq a^{1/2} + 1 - ca^{-1}$$

(for these special values of a).

If we tried to define an impact volume in the continuous case, we would recover the volume, at least for compact sets. Still, the above results and questions suggest that ordinary volume is not the best tool to understand additive properties. Perhaps one could try to modify the definition of impact volume by requiring $\mu(A) \geq \mu(B)$. So put

$$\operatorname{iv}_*(B) = \inf_{a \geq \mu(B)} \left(\xi_B(a)^{1/d} - a^{1/d}\right)^d.$$

PROBLEM 4.7. Find a lower estimate for $\operatorname{iv}_*(B)$ in terms of $\mu(B)$ and $\mu(\operatorname{conv} B)$.

5. Topologies on integers

For most of the time we are happy with the integers as a discrete set. However, other important topologies do exist on them. An example is the p-adic topology, which can be compactified to give us p-adic integers. The Čech-Stone compactification also has applications in combinatorial number theory.

In the sequel we will always use commutative groups ($\mathbb{Z}$ and extensions); for non-commutative groups the following definitions and claims have to be modified a litle.

Among all topologies we will be interested in those where addition behaves nice. We consider two possible interpretations, a stronger and a weaker one.

DEFINITION 5.1. Let G be a group and $\mathcal{T}$ a topology on it. We say that $(G, \mathcal{T})$ is a *topological group*, if addition and substraction are continuous in $\mathcal{T}$, that is, $f(x, y) = x - y$ is jointly continuous in both variables. It is a *semitopological group*, if $x - y$ is continuous in each variable separately.

The weaker condition means two things: first, if U is a neighbourhood of 0, then so is $-U$; next, if U is a neighbourhood of an element x, then $U + a$ is a neighbourhood of $x + a$. The stronger condition, in addition, requires that for any neighbourhood U of 0 there is another neighbourhood U' such that $U' - U' \subset U$.

EXERCISE 84. Construct a topology on $\mathbb{Z}$ which makes it a semitopological group but not a topological group.

When defining a topology on $\mathbb{Z}$, we shall typically proceed as follows. We define a basis of neigbourhood of 0. Then, to make $\mathbb{Z}$ at least a semitopological group, we define neighbourhoods of other integers by translation, that is, U will be a neighbourhood of x if $U - x$ is a neighbourhood of 0. In most cases it will be trivial that this indeed defines a topology, and we shall not give the easy details. If there is a hidden difficulty, we shall point it out.

A topological group may be *complete*, which means that every sequence (assuming a countability condition) or every generalized sequence (indexed by a general ordered set, not necessarily by positive integers) which satisfies the Cauchy condition must be convergent.

A topological group can always be *completed* by assigning a new element (a limit) to every (generalized) Cauchy sequence which does not already have one. This procedure is used to build p-adic integers.

If we are lucky, this completion is compact and then we can embed our group into a compact group, that is, we can *compactify* it.

EXERCISE 85. Construct a group topology on $\mathbb{Z}$, other than the discrete one, which cannot be compactified.

We can sometimes compare two topologies as follows.

DEFINITION 5.2. Let $\mathcal{T}$, $\mathcal{T}'$ be two topologies on the same set. We say that $\mathcal{T}'$ is *finer* than $\mathcal{T}$, or $\mathcal{T}$ is *coarser* than $\mathcal{T}'$, if every set which is open in $\mathcal{T}$ is also open in $\mathcal{T}'$.

The finest topology of all is the discrete one. The coarsest is the one in which only the empty set and the whole space are open, a pretty uninteresting one.

In the sequel we will find the answer to the following questions.

Question 1. What is the coarsest topology on $\mathbb{Z}$ in which all characters are continuous?

Recall that a character is a homomorphism into the circle $\{z \in \mathbb{C} : |z| = 1\}$. Now if γ is a character and $\gamma(1) = \omega = e^{2\pi i t}$, then necessarily

$$\gamma(n) = \omega^n = e^{2\pi i t n}$$

for all integers.

Question 2. What is the finest topology on $\mathbb{Z}$ which can be compactified to make it a compact topological group?

We shall see that the answer to these questions will be the same, and we shall call it the *Bohr topology*.

We first answer the first question.

By the multiplicative property of characters continuity everywhere is equivalent to continuity at 0. This means that the following sets must be neighbourhoods of 0 for each character γ and $\varepsilon > 0$:

$$\{n \in \mathbb{Z} : |\gamma(n) - \gamma(0)| < \varepsilon\}.$$

If we exptess γ as $\gamma(n) = e^{2\pi i t n}$, we see that this set is the same as

$$\{n \in \mathbb{Z} : \|tn\| < \delta\},$$

where ε and δ are connected by the equation

$$\varepsilon = \left|e^{2\pi i \delta} - 1\right| = 2 \sin \pi\delta.$$

If such a set has to be a neighbourhood of 0, then so has any finite intersection of such sets. We saw similar objects in Section 5 of Chapter 2, which we now repeat.

Definition 5.3. If G is a commutative group, $\gamma_1, \dots, \gamma_k$ are characters of G and $\varepsilon_j > 0$, we write

$$B(\gamma_1, \dots, \gamma_k; \varepsilon_1, \dots, \varepsilon_k) = \{g \in G : |\arg \gamma_j(g)| < 2\pi\varepsilon_j \text{ for } j = 1, \dots, k\}$$

and call these sets *Bohr sets*. In particular, if $\varepsilon_1 = \dots = \varepsilon_k = \varepsilon$, we shall speak of a *Bohr* (k, ε)*-set*. (We take the branch of arg that lies in $[-\pi, \pi)$.)

By view of the above, a Bohr set in $\mathbb{Z}$ can also be written as

$$B(u_1, \dots, u_k; \varepsilon_1, \dots, \varepsilon_k) = \{x \in \mathbb{Z} : \|u_j x\| < \varepsilon_j \text{ for } j = 1, \dots, k\},$$

where now the parameters u_j are real numbers taken modulo 1.

Definition 5.4. The *Bohr topology* on a commutative group is the topology in which a set is a neigbourhood of a point x if and only if it contains a set of the form $x + N$, where B is a Bohr set.

Exercise 86. The Bohr sets are open in the Bohr topology.

Exercise 87. The Bohr topology turns $\mathbb{Z}$ into a topological group.

Exercise 88. This group can be compactified.

Exercise 89. A sequence is convergent in the Bohr topology only if it is constant from a point on.

6. The finest compactification

Now we answer Question 2 of the previous section.

Let G be a compact group, and let U be a neighbourhood of 0 in G. The collection of open sets

$$\{U + x : x \in G\}$$

covers G, hence so does a finite subcollection. Hence U has the property that there are finitely many elements $x_1, \dots, x_k \in G$ such that

$$\bigcup (U + x_i) = G.$$

Definition 6.1. A set A in a group G is *syndetic*, if there are finitely many elements $x_1, \dots, x_k \in G$ such that

$$\bigcup (A + x_i) = G.$$

So a neighbourhood of 0 must be syndetic.

EXERCISE 90. A set $A \subset \mathbb{Z}$ is syndetic if and only if it is unbounded both from above and from below, and has bounded gaps, that is,

$$A = \{\dots, a_{-1}, a_0, a_1, a_2, \dots\},\ a_{k+1} - a_k < c$$

with some c.

EXERCISE 91. Bohr sets are syndetic.

Let U be a neighbourhood of zero in any conditionally compact topology (this expression means that it has a compactification). We can find an open set U_1 such that $U_1 - U_1 \subset U$. Then, there is an open set U_2 such that $U_2 - U_2 \subset U_1$, and so on. So, U can be a neighbourhood of zero only if there is a chain of syndetic sets $U_1, U_2, \dots$ such that $U_{i+1} - U_{i+1} \subset U_i \quad \forall i$.

Note that the Bohr sets have this property, because

$$B(\alpha_1, \dots, \alpha_k, \varepsilon/2) - B(\alpha_1, \dots, \alpha_k, \varepsilon/2) \subset B(\alpha_1, \dots, \alpha_k, \varepsilon).$$

It is possible to show directly (without any reference to characters) that this requirement defines a class of sets which can serve as the basis of a topology, and it is indeed conditionally compact. Instead we shall prove that any set with this property contains a Bohr set. Moreover, we do not need an infinite chain for this, two steps are sufficient.

Note that for a sequence $U, U_1, U_2, \dots$ of sets $U_1, U_2, \dots$ such that $U_{i+1} - U_{i+1} \subset U_i$ we have

$$(U_2 - U_2) - (U_2 - U_2) = 2U_2 - 2U_2 \subset U.$$

So the exact formulation of the above claim sounds as follows.

THEOREM 6.2 (Bogolyubov). *If $A \subset \mathbb{Z}$ is a syndetic set, then $2A - 2A$ contains a Bohr set (in other words, it is a neighbourhood of 0 in the Bohr topology).*

This will be proved in a stronger form in the next section.

7. Banach density

To put Bogolyubov's theorem into proper perspective we define some new concepts of density.

DEFINITION 7.1. The *lower* and *upper Banach densities* of a set A of integers are defined by

$$d_*(A) = \lim_{n\to\infty} \min_x \frac{|A \cup [x+1, x+n]|}{n}$$

$$d^*(A) = \lim_{n\to\infty} \max_x \frac{|A \cup [x+1, x+n]|}{n}$$

EXERCISE 92. These limits do exist.

EXERCISE 93. Show that for any set $A \subset \mathbb{Z}$, $d_*(A) \le \underline{d}(A) \le \overline{d}(A) \le d^*(A)$.

EXERCISE 94. $d_*(A) > 0$ if and only if A is syndetic.

EXERCISE 95. Let P be the set of primes. Show that $d^*(P) = 0$.

The stronger form of Bogolyubov's theorem requires only positive upper Banach density.

THEOREM 7.2 (Bogolyubov). *If $d^*(A) > 0$, then there exist $\alpha_1, \dots, \alpha_k$, and $\varepsilon > 0$, with k and ε depending only on $d^*(A) > 0$, such that $B(\alpha_1, \dots, \alpha_k, \varepsilon) \subset 2A - 2A$*

PROOF. We split the proof into three steps.

Step 1, modular case: It corresponds to Lemma 5.2, which we repeat below.

LEMMA 7.3. *Let G be a finite commutative group, $|G| = q$. Let A be a nonempty subset of G and write $|A| = m = \beta q$. The set $D = 2A - 2A$ (the second difference set of A) contains a Bohr (k, ε)-set with some integer $k < \beta^{-2}$ and $\varepsilon = 1/4$.*

Step 2, finite case:

LEMMA 7.4. *If $A \subset [t, t + l]$, $|A| \geq \beta l$, then there are $\alpha_1, \dots, \alpha_k, \varepsilon$, with k, ε depending only on β such that $2A - 2A \supseteq B(\alpha_1, \dots, \alpha_k, \varepsilon) \cap [-l, l]$.*

To see this let $q > 4l$ and let $A \subset [t, t + l]$. We denote by $A' \subset \mathbb{Z}_q$ the set of residue classes of A modulo q, which satisfies $|A'| \geq \beta' l$ with $\beta' = \beta/5$. Then, by the modular case, there exist $k(\beta), u_1, \dots, u_k$ and $\varepsilon(\beta)$ such that

$$2A' - 2A' \supseteq \left\{ x : \left\| \frac{xu_i}{q} \right\| < \varepsilon \right\}.$$

This means that for any n such that $\|n\alpha_i\| < \varepsilon$, where $\alpha_i = u_i/q$, we can find $a_1, a_2, a_3, a_4 \in A$ such that $n \equiv a_1 + a_2 - a_3 - a_4 mod q$ with $|a_1 + a_2 - a_3 - a_4| < 2l$. If $|n| < 2l$, they only can be congruent if they are equal. So, for all $n \in B(\alpha_1, \dots, \alpha_k, \varepsilon) \cap [-l, l]$ we have $n = a_1 + a_2 - a_3 - a_4$, which completes the second step of the proof.

Last step, density case: Let β satisfy $d^*(A) > \beta > 0$. Given $l \in \mathbb{N}$, we can find t such that $|A \cup [t, t + l]| \geq \beta l$. The finite case provides a finite collection $\alpha_1, \dots, \alpha_k, \varepsilon$ such that $2A - 2A \supseteq B(\alpha_1, \dots, \alpha_k, \varepsilon) \cap [-l, l]$. Note that k and ε are fixed, while $\alpha_1, \dots, \alpha_k$, which we can assume to belong to $[0, 1]$, depend on l.

For each l, we define

$$C_l = \{(\alpha_1, \dots, \alpha_k) \in [0, 1]^k \ : \ B(\alpha_1, \dots, \alpha_k, \varepsilon) \cap [-l, l] \subseteq 2A - 2A\}$$

For all l, C_l is a compact set, nonempty by the previous argument, and $C_{l+1} \subset C_l$. Then,

$$\bigcap_{l \in \mathbb{N}} C_l \neq \emptyset.$$

So there is at least one element $(\alpha_1, \dots, \alpha_k)$ in the intersection. This defines the Bohr neigbourhood in $2A - 2A$ we were looking for.

□

8. The difference set topology

By Bogolyubov's theorem, with 4 copies of A we get a Bohr neighbourhood of zero, namely by forming $A + A - A - A$. With some modification 3 copies also suffice. Denote

$$k \cdot A = \{ka, a \in A\}.$$

THEOREM 8.1 (Bergelson-Ruzsa [**1**]). *Let A be a set of integers with $d^*(A) > 0$, and let r, s, t be integers such that $r + s + t = 0$. The set $r \cdot A + r \cdot A + t \cdot A$ is a Bohr neighbourhood of zero. In particular, the set $2A - 2 \cdot A = A + A - 2 \cdot A$ is a Bohr neighbourhood of zero.*

The question whether 2 copies of A is enough to have a Bohr neighbourhood inside is difficult, and the answer may depend on the interpretation of density.

Unsolved problem: If A has positive lower Banach density ($d_*(A) > 0$), is $A - A$ a Bohr neighbourhood of 0?

Here we used the strongest assumption with lower Banach density, and in Bogolyubov's theorem the weakest one with upper Banach density was sufficient. The medium ones with asymptotic density do not give anything new, for problems about the diffence set they behave as lower Banach density. An exact formulation is as follows.

THEOREM 8.2 ([**42**]). *Assume $d^*(A) > 0$. Then there is an $A' \subset \mathbb{N}$ such that $d(A') > 0$ and $A' - A' \subset A - A$.*

However, upper Banach density is different.

THEOREM 8.3 (Kříž[**28**]). *There exist an A with $d(A) > 0$ such that there is no A', with $d_*(A) > 0$ and $A' - A' \subset A - A$. Consequently, $A - A$ is not a Bohr neighbourhood of 0.*

The following result guarantees a somewhat weaker property.

THEOREM 8.4 (Følner[**7, 8**]). *If $d^*(A) > 0$, then exist a $B = B(\alpha_1, \dots, \alpha_k, \epsilon)$ such that $d((A - A) \setminus B) = 0$.*

Unsolved problem: If A has positive upper Banach density, is $A - A$ a neighbourhood of something?

So we know that difference sets of sets of positive density are not necessarily neighbourhoods of 0 in the Bohr topology. They are, however, neighbourhoods in some topology: we can use them to define a new topology.

DEFINITION 8.5. We say that $V \subset \mathbb{Z}$ is a neighbourhood of 0 in the *difference set topology* if there exist a set A with $d^*(A) > 0$ such that $A - A \subset V$. V' is said to be a neighbourhood of $n \in \mathbb{Z}$ if $V = V' - n$ is a neighbourhood of 0.

EXERCISE 96. The difference set topology is indeed a topology (not easy!).

Easier exercise: formulate what this means using only sets and density, no topological concept.

EXERCISE 97. Is the difference set topology a group topology?

DEFINITION 8.6. The *syndetic difference topology* is defined similarly to the difference set topology, but not we saiy that $V \subset \mathbb{Z}$ is a neighbourhood of zero if there exist an A with $d_*(A) > 0$ such that $A - A \subset V$.

With these concepts we can reformulate Kříž' theorem as the syndetic difference topology is different from the difference set topology, and we cannot decide whether it is the same as the Bohr topology.

DEFINITION 8.7. The *combinatorial difference topology* is defined as follows. Let $A_1, \dots A_k$ subsets of the integers such that $\mathbb{Z} = \bigcup_{i=1}^{n} A_i$, then $\bigcup_i (A_i - A_i)$ is a neighbourhood of 0.

EXERCISE 98. The syndetic difference topology and the combinatorial difference topology are, indeed, topological spaces.

EXERCISE 99. If V is an open set in the syndetic difference topology so is in the combinatorial. Hence the syndetic difference topology and the combinatorial difference topology are identical.

Bibliography

1. V. Bergelson and I. Z. Ruzsa, *Sumsets in difference sets*, Israel J. Math.
2. Y. Bilu, *Structure of sets with small sumset*, Structure theory of set addition, Astérisque, vol. 258, Soc. Mat. France, 1999, pp. 77–108.
3. N. N. Bogolyubov, *Some algebraical properties of almost periods*, Zap. Kafedry Mat. Fiziki Kiev **4** (1939), 185–194.
4. J. Bourgain, *On triples in arithmetic progression*, GAFA **9** (1999), 968–984.
5. Mei-Chu Chang, *A polynomial bound in Freiman's theorem*, (preprint).
6. György Elekes and Zoltán Király, *On combinatorics of projective mappings*, **14** (2001), 183–197.
7. E. Følner, *Generalization of a theorem of Bogoliuboff to topological Abelian groups. with an appendix on Banach mean values in non-Abelian groups*, Math. Scandinavica **2** (1954), 5–18.
8. _____, *Note on a generalization of a theorem of Bogoliuboff*, Math. Scandinavica **2** (1954), 224–226.
9. G. Freiman, *Inverse problems in additive number theory vi. on the addition of finite sets iii (in russian)*, Izv. Vyss. Ucebn. Zaved. Matematika **3 (28)** (1962), 151–157.
10. _____, *Foundations of a structural theory of set addition*, American Math. Soc., 1973.
11. G. Freiman and V. P. Pigaev, *The relation between the invariants r and t (russian)*, Kalinin. Gos. Univ. Moscow (1973), 172–174.
12. R. J. Gardner and P. Gronchi, *A Brunn-Minkowski inequality for the integer lattice*, Trans. Amer. Math. Soc. **353** (2001), 3995–4024.
13. W. T. Gowers, *A new proof of Szemerédi's theorem for arithmetic progressions of length four*, Geom. Funct. Anal. **8** (1998), 529–551.
14. _____, *A new proof of Szemerédi's theorem*, Geom. Funct. Anal. **11** (2001), 465–88.
15. A. Granville, *An introduction to additive combinatorics*, Additive Combinatorics (Providence, RI, USA), CRM Proceedings and Lecture Notes, vol. 43, American Math. Soc., 2007, pp. 1–27.
16. B. J. Green and I. Z. Ruzsa, *Freiman's theorem in an arbitrary Abelian group*, J. London Math. Soc. **75** (2007), 163–175.
17. K. Gyarmati, S. Konyagin, and I. Z. Ruzsa, *Double and triple sums modulo a prime*, Additive Combinatorics (Providence, RI, USA) (A. Granville, M.B. Nathanson, and J. Solymosi, eds.), CRM Proceedings and Lecture Notes, vol. 43, American Math. Soc., 2007, pp. 271–277.
18. Katalin Gyarmati, M. Matolcsi, and I. Z. Ruzsa, *Plünnecke's inequality for different summands*, in preparation.
19. _____, *A superadditivity and submultiplicativity property for cardinalities of sumsets*, Combinatorica, to appear.
20. H. Halberstam and H.; K. F. Roth, *Sequences*, Clarendon, London, 1966, 2nd ed. Springer, 1983.
21. P. Hegedűs, G. Piroska, and I. Z. Ruzsa, *On the Schnirelmann density of sumsets*, Publ. Math. Debrecen **53** (1998), 333–345.
22. F. Hennecart, G. Robert, and A. Yudin, *On the number of sums and differences*, Structure theory of set addition, Astérisque, vol. 258, Soc. Mat. France, 1999, pp. 173–178.
23. J. H. B. Kemperman, *On products of sets in a locally compact group*, Fundamenta Math. **56** (1964), 51–68.
24. A. G. Khovanskii, *Newton polyhedron, Hilbert polynomial, and sums of finite sets*, Functional Anal. Appl. **26** (1992), 276–281.
25. _____, *Sums of finite sets, orbits of commutative semigroups, and hilbert functions*, Functional Anal. Appl. **29** (1995), 102–112.

26. M. Kneser, *Abschätzungen der asymptotischen Dichte von Summenmengen*, Math. Zeitschrift **58** (1953), 459–484.
27. ———, *Summenmengen in lokalkompakten Abelschen Gruppen*, Math. Zeitschrift **66** (1956), 88–110.
28. I. Kříž, *Large independent sets in shift-invariant graphs*, Graphs and Comb. **3** (1987), 145–158.
29. B. Lepson, *Certain best possible results in the theory of Schnirelmann density*, Proc. Amer. Math. Soc. **1** (1950), 592–594.
30. V. F. Lev and P. Smeliansky, *On addition of two distinct sets of integers*, Acta Arithmetica **70** (1995), 85–91.
31. A. M. Macbeath, *On measure of sum sets II.*, Proc. Cambridge Phil. Soc. **49** (1953), 40–43.
32. J. L. Malouf, *On a theorem of Plünnecke concerning the sum of a basis and a set of positive density*, J. Number Theory **54**.
33. M. B. Nathanson, *Additive number theory: Inverse problems and the geometry of sumsets*, Springer, 1996.
34. ———, *Growth of sumsets in abelian semigroups*, Semigroup Forum **61** (2000), 149–153.
35. M. B. Nathanson and I. Z. Ruzsa, *Polynomial growth of sumsets in abelian semigroups*, J. Th. Nombres Bordeaux **14** (2002), 553–560.
36. Oystein Ore, *Theory of graphs*, Amer. Math. Soc., Providence, RI, USA, 1962 (reprints 1967, 1974.
37. H. Plünnecke, *Über die dichte der summe zweier mangen, deren eine die dichte null hat*, J. Reine Angew. Math. **205** (1960), 1–20.
38. ———, *Eigenschaften und Abschätzungen von Wirkungsfunktionen*, Gesellschaft für Mathematik und Darenverarbeitung, Bonn, 1969.
39. ———, *Eine zahlentheoretische anwendung der graphtheorie*, J. Reine Angew. Math. **243** (1970), 171–183.
40. D. A. Raikov, *On the addition of point sets in the sense of schnirelmann*, Mat. Sbornik **5**.
41. C. A. Rogers and G. C. Shephard, *The difference body of a convex body*, Arch. Math **8** (1957), 220–233.
42. I. Z. Ruzsa, *On difference sets*, Studia Sci. Math. Hungar. **13** (1978), 319–326.
43. ———, *On the cardinality of $A+A$ and $A-A$*, Combinatorics (Keszthely 1976), Coll. Math. Soc. J. Bolyai, vol. 18, North-Holland – Bolyai Társulat, Budapest, 1978, pp. 933–938.
44. ———, *Essential components*, Proc. London Math. Soc. **54** (1987), 38–56.
45. ———, *An additive property of squares and primes*, Acta Arithmetica **49** (1988), 281–289.
46. ———, *An additive problem for powers of primes*, J. Number Theory **33** (1989), 71–82.
47. ———, *An application of graph theory to additive number theory*, Scientia, Ser. A **3** (1989), 97–109.
48. ———, *Addendum to: An application of graph theory to additive number theory*, Scientia, Ser. A **4** (1990/91), 93–94.
49. ———, *Diameter of sets and measure of sumsets*, Monats. Math. **112** (1991), 323–328.
50. ———, *Arithmetical progressions and the number of sums*, Periodica Math. Hung. **25** (1992), 105–111.
51. ———, *A concavity property for the measure of product sets in groups*, Fundam. Math. **140** (1992), 247–254.
52. ———, *On the number of sums and differences*, Acta Math. Sci. Hungar **59** (1992), 439–447.
53. ———, *Generalized arithmetical progressions and sumsets*, Acta Math. Hung. **65** (1994), 379–388.
54. ———, *Sum of sets in several dimensions*, Combinatorica **14** (1994), 485–490.
55. ———, *Sets of sums and commutative graphs*, Studia Sci. Math. Hungar. **30** (1995), 127–148, Proc. of the workshop in combinatorics, Bielefeld 1991.
56. ———, *Sums of finite sets*, Number Theory, New York seminar 1991-1995 (D. V. Chudnovsky, G. V. Chudnovsky, and M. B. Nathanson, eds.), Springer–Verlag, New York–Berlin, 1996, pp. 281–293.
57. ———, *The Brunn-Minkowski inequality and nonconvex sets*, Geometricae Dedicata **67** (1997), 337–348.
58. A. Schields, *Sur la mesure d'une somme vectorielle*, Fundamenta Math. **42** (1955), 57–60.
59. E. Szemerédi, *On sets of integers containing no k elements in arithmetic progression*, Acta Arithmetica **27** (1975), 299–345.