Introduction to the Theory of Set Addition

October 6th – 10th 2014, Freie Universität Berlin

What is set addition and why should one learn the basics of this theory?

The first question is easy to answer. The word 'addition' is mentioned so there must be an ambient commutative group. The word 'set' is mentioned so we must add sets.

The *sumset* or *Minkowski sum* of two sets A and B in a commutative group is defined to be

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

Adding a singleton to another set is translation so we will use the standard notation

$$\{a\} + B = a + B.$$

The iterated sumset h-fold hA is defined recursively by

$$hA = (h-1)A + A.$$

The *difference set* naturally is

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

By $kA - \ell A$ we mean the set

$$kA - \ell B = \{a_1 + \cdots + a_k - b_1 - \cdots - b_\ell : a_j \in A, b_i \in B\}.$$

The second questions is trickier. To begin to answer it, let us introduce a concept you may actually not have heard before: that of a set of small doubling. The *doubling constant* of a finite set A is the ratio

$$\text{doubling constant of } A := \frac{|A + A|}{|A|}.$$

One can think of the doubling constant as a measure of "additive structure". Finite subgroups, which are closed under addition and in general have very rich structure, have doubling one, while a finite set A of generators of a free commutative group has maximum doubling $\binom{|A|+1}{2}$.

The doubling constant also allows one to study questions that are otherwise almost meaningless. For example, given a finite non-empty set A in a commutative group, what can be said about the cardinality of $A+A+A$? Well, not much. The bounds

$$|A| \leq |A+A+A| \leq \binom{|A|+2}{3}$$

are easy to prove and sharp: the lower bound is attained when A is a subgroup and the upper bound when A is a set of generators of a free commutative group.

Once a condition on the doubling constant is inserted, the question becomes more precise, more difficult to answer and also more useful. The archetypical question we will study in the first couple of days is as follows.

> How large can $|A+A+A|$ be, when $|A+A| \leq \alpha|A|$?

Later on in the Block Course you will see a rather precise characterisation of sets of small doubling. This celebrated theorem of Freiman has influenced additive number theory in the 21st century considerably. As you will see, many of the basic techniques and results we will learn this week are featured in the study of sets of small doubling.

They also appear in many famous results of famous mathematicians: Ruzsa's proof of Freiman's theorem, Gowers' proof of Szemerédi's theorem, Bourgain's contribution to the Kakeya problem, the Bourgain-Katz-Tao sum-product theorem for finite fields and Helfgott's result about growth and generation in $SL_2(\mathbb{Z}/p\mathbb{Z})$.

Moreover, there are sound educational reasons for studying this subject. Many particularly useful combinatorial techniques are applied: double-counting, working with extreme quantities, probabilistic reasoning and the Cauchy-Schartz inequality. I hope that in this first week you will see how one can do a lot by using very little.

Acknowledgement. In preparing these handouts, I relied on various material of Ben Green, Tim Gowers and Imre Ruzsa.

Please email corrections to g.petridis@rochester.edu

1 Cardinality inequalities

Let us begin with a gentle exercise.

Real time exercise. For each of the following sets determine the desired quantities.

(i) $A = \{1, \dots, n\} \subset \mathbb{Z}$. Find $A - A$.

(ii) $A = \{1, \dots, n\} \times \{1, \dots, n\} \subset \mathbb{Z}^2$. Find $A + A$ and the doubling constant.

(iii) $A = \{\mathbf{e}_1, \dots, \mathbf{e}_d\} \subset \mathbb{R}^d$. Find $A + A$ and the doubling constant. $\{\mathbf{e}_i\}$ is the standard basis.

Solution.

From now on all sets are finite, non-empty subsets of a commutative group

Our first topic is to study what the doubling constant of a finite set tells us about the cardinality of sum-and-difference sets like the ones defined above. We begin with a remarkable inequality of Ruzsa.

Lemma 1.1 (Ruzsa's triangle inequality)**.** *Let X, Y, Z be finite non-empty sets in a commutative group. Then*

$$|X||Y - Z| \leq |Y - X||X - Z|.$$

Sketch of proof. We construct an injection from $X \times (Y - Z) \mapsto (Y - X) \times (X - Z)$.

Let $(x, v) \in X \times (Y - Z)$. Express $v = y - z$ for some $y \in Y$ and $z \in Z$. Then map $(x, y - z)$ to $(y - x, x - z)$.

Corollary 1.2 (From sums to differences)**.** *Let $\alpha \in \mathbb{R}$ and A be a finite no-empty set in a commutative group. Suppose that $|A + A| \leq \alpha|A|$. Then $|A - A| \leq \alpha^2|A|$.*

Proof. The strategy is to "insert an additional A between the two copies of A" in $A-A$.

Set $X=-A$, $Y=Z=A$ in the triangle inequality. Then $|-A||A-A| \le |A-(-A)||(-A)-A|$. This implies $|A||A-A| \le |A+A|^2$ and so

$$|A-A| \le \frac{|A+A|}{|A|}|A+A| \le \alpha|A+A| \le \alpha^2|A|.$$

□

This simple lemma is a remarkable result: it is easy to state, easy to prove, essentially sharp and with many applications!

The exponent of α is sharp. There are examples of arbitrarily large values of α where

$$|A+A| = \alpha|A| \text{ and } |A-A| \ge c\frac{\alpha^2}{\sqrt{\log(\alpha)}}|A|,$$

for some absolute constant c. Absolute means a genuine constant: independent of A (and hence of α).

Application 1.3 (From three to many summands)**.** *Let $\beta \in \mathbb{R}$ and A be a finite no-empty set in a commutative group. Suppose that $|A+A-A| \le \beta|A|$. Then $|A+A+A+A| \le \beta^2|A|$.*

Proof. The strategy is to "insert an A between the two copies of $A+A$" in $4A$.

Set $X=A$, $Y=A+A$ and $Z=-(A+A)$ in the triangle inequality. Then $|-A||A+A+A+A| \le |A+A-A||A-(A+A)|$. This implies

$$|A+A+A+A| \le \frac{|A+A-A|}{|A|}|A+A-A| \le \beta^2|A|.$$

□

Bounds on $\lvert A \pm A \pm A\rvert$ imply bounds on $\lvert kA-\ell A\rvert$

Real time exercise. Let $\alpha \in \mathbb{R}$ and A be a finite no-empty set in a commutative group. Suppose that $|A+A| \le \alpha|A|$. What bound can you put on $|A+A+A|$?

Solution.

Lemma 1.4 (From two to many summands)**.** *Let A, B be finite non-empty sets in a commutative group. Let $\emptyset \neq X \subseteq A$ be a non-empty subset of A that minimises the quantity $|Z+B|/|Z|$ over all non-empty subsets of A. Then for all non-empty sets C in the ambient group*

$$|X||X+B+C| \leq |X+B||X+C|.$$

Proof. For simplicity let us denote

$$K := \frac{|X+B|}{|X|} = \min_{\emptyset \neq Z \subseteq A} \frac{|Z+B|}{|Z|}.$$

A is finite so the minimum (exists and) is attained. The inequality in the statement of the lemma now becomes

$$|X+B+C| \leq K|X+C|.$$

To prove it, we induct on $|C|$.

When $|C| = 1$ we are done as both sides of the inequality equal $|X+B|$.

For $|C| > 1$ we pick $c \in C$ and express $C = C' \cup \{c\}$. We now partition X into a set T and its relative complement $X \setminus T$, where

$$t \in T \iff c + t \in C' + X.$$

Then

$$\begin{aligned} |X+B+C| &= |X+B+C'| + |(X+B+c) \setminus (X+B+C')| \\ &\leq |X+B+C'| + |(X+B+c) \setminus (T+B+c)| \\ &= |X+B+C'| + |X+B+c| - |T+B+c| \\ &= |X+B+C'| + |X+B| - |T+B|. \end{aligned}$$

By the induction hypothesis the first summand is at most $K|X+C'|$. By the definition of K the second summand is $K|X|$. By the definition of X, $|T+B| \geq K|T|$. Therefore

$$|X+B+C| \leq K(|X+C'| + |X| - |T|) = K(|X+C'| + |X \setminus T|).$$

This completes the proof as T is such that $|X+C| = |X+C'| + |X \setminus T|$. □

At the heart of the proof of the lemma lies *submodularity*: for any two sets S and T we have

$$|X + (S \cup T)| + |X + (S \cap T)| \leq |X+S| + |X+T|.$$

To verify the above inequality note

$$|X+S|+|X+T| = |(X+S)\cup(X+T)|+|(X+S)\cap(X+T)| \geq |X+(S\cup T)|+|X+(S\cap T)|.$$

It is noteworthy that the Cauchy–Davenport and Kneser inequalities, which offer basic lower bounds on the cardinality of sumsets, are also related to submodularity.

Theorem 1.5 (Plünnecke's inequality)**.** *Let $\alpha \in \mathbb{R}$ and A and B be finite non-empty sets in a commutative group. Suppose that $|A+B| \leq \alpha|A|$. There exists a non-empty subset $\emptyset \neq X \subseteq A$ such that for all positive integers h*

$$|X + hB| \leq \alpha^h |X|.$$

In particular, if $|A + A| \leq \alpha|A|$, then $|hA| \leq \alpha^h|A|$.

Proof. As in the proof of Lemma 1.4, we let X and K be such that

$$K := \frac{|X+B|}{|X|} = \min_{\emptyset \neq Z \subseteq A} \frac{|Z+B|}{|Z|}.$$

Setting $Z = A$ yields $K \leq \alpha$.

We now induct on h.

For $h = 1$ note $|X + B| = K|X| \leq \alpha|X|$.

For $h > 1$ we set $C = (h-1)B$ in Lemma 1.4:

$$|X + hB| = |X + B + (h-1)B| \leq K|X + (h-1)B| \leq K^h|X| \leq \alpha^h|X|.$$

For the second inequality we set $B = A$ and use the fact that X is non-empty.

$$|hA| \leq |X + hA| \leq \alpha^h|X| \leq \alpha|A|.$$

□

Remarks. In general it is not possible to replace X by the whole of A. Note the order of the quantifiers: there is an X that works for all h.

Theorem 1.6 (The Plünnecke–Ruzsa inequalities)**.** *Let $\alpha \in \mathbb{R}$, k and ℓ be positive integers, and A be a finite non-empty set in a commutative group. Suppose that $|A + A| \leq \alpha|A|$. Then*

$$|kA - \ell A| \leq \alpha^{k+\ell}|A|.$$

Proof. We combine Ruzsa's triangle inequality (Lemma 1.1) with Plünencke's inequality Theorem 1.5). We use the fact that the same set X works for both k and ℓ.

$$|kA - \ell A| \leq \frac{|X + kA||X + \ell A|}{|X|} \leq \frac{\alpha^k |X| \alpha^\ell |X|}{|X|} = \alpha^{k+\ell}|X|.$$

□

Bounds on $|A \pm A|$ imply bounds on $|kA - \ell A|$

Lemma 1.7 (Ruzsa's twin to the triangle inequality)**.** *Let A, B, C be finite non-empty sets in a commutative group. Then*

$$|A||B + C| \leq |A + B||A + C|.$$

Proof. Once again we apply Lemma 1.4, noting that $X \neq \emptyset$ and $\dfrac{|X + B|}{|X|} \leq \dfrac{|A + B|}{|A|}$.

$$\begin{aligned} |A||B + C| &\leq |A||X + B + C| \\ &\leq |A|\frac{|X + B|}{|X|}|X + C| \\ &\leq |A|\frac{|A + B|}{|A|}|A + C| \\ &= |A + B||A + C|. \end{aligned}$$

□

Lemma 1.8 (Another cardinality inequality of Ruzsa)**.** *Let A, B, C be finite non-empty sets in a commutative group. Then*

$$|A + B + C|^2 \leq |A + B||B + C||C + A|.$$

No proof given here. Note that the above inequality bounds the whole of $|A + B + C|$ and not just $|X + B + C|$ for some suitably chosen X.

Let us quickly compare the range of α for which the bound above beats that of Plünnecke's inequality. Setting $A = B = C$ yields

$$|A + A + A| \leq \alpha^{3/2}|A|^{3/2}.$$

This is superior to $\alpha^3|A|$ when $|A| \leq \alpha^3$ or $\alpha \geq |A|^{1/3}$.

2 The power trick

A very natural question is to ponder what happens when we add different sets to A. As one may expect the outlook does not change much.

Theorem 2.1 (Ruzsa's Plünnecke–type inequality for different summands)**.** *Let h be a positive integer and $A, B_1, \dots, B_h$ be finite non-empty sets in a commutative group. Suppose that $|A + B_i| \le \alpha|A|$ for all $i = 1, \dots, h$. Then*

$$|B_1 + \cdots + B_h| \le \alpha^h |X|.$$

Proof. We apply Plünnecke's inequality (Theorem 1.5) to the sets A and $B_1 \cup \cdots \cup B_h$. Note

$$|A + (B_1 \cup \cdots \cup B_h)| = |(A + B_1) \cup \cdots \cup (A + B_h)| \le |A + B_1| + \cdots + |A + B_h| \le h\alpha|A|.$$

By Theorem 1.5 we get a non-empty $\emptyset \neq X \subseteq A$ such that

$$|X + h(B_1 \cup \cdots \cup B_h)| \le (h\alpha)^h |X| \le h^h \alpha^h |A|. \tag{2.1}$$

Now observe

$$\begin{aligned} |B_1 + \cdots + B_h| &\le |X + B_1 + \cdots + B_h| \\ &\le |X + h(B_1 \cup \cdots \cup B_h)| \\ &\le h^h \alpha^h |A|. \end{aligned}$$

The second step is to find a way to remove the h^h term.

To do this we use the fact that inequality (2.1) holds for all sets in any commutative group. We work in the r-fold direct product of the ambient group and consider the r-fold Cartesian products $A^r = A \times \ldots \times A$ and $B_i^r = B_i \times \ldots \times B_i$.

There are two key observations to be made.

(i) Cartesian products and set addition mix well together. For example,

$$B_1^r + \cdots + B_h^r = (B_1 + \cdots + B_h)^r.$$

(ii) Cardinality is *multiplicative* with respect to cartesian products. For example

$$|A^r + B_i^r| = |(A + B_i)^r| = |A + B_i|^r \le \alpha^r |A|^r = \alpha^r |A^r|.$$

More informally working in the r-fold direct product results in:

- $|A + B_1 + \cdots + B_h|$ being replaced by its rth power,

- α being replaced by its rth power,

- $|A|$ being replaced by its rth power,

- h^h remaining as it is.

Applying inequality (2.1) to the sets $A^r, B_1^r \ldots, B_h^r$ gives

$$|B_1 + \cdots + B_h|^r = |B_1^r + \cdots + B_h^r| \leq h^h(\alpha^r)^h|A^r| = h^h(\alpha^h)^r|A|^r.$$

Taking rth roots gives

$$|B_1 + \cdots + B_h| \leq (h^h)^{1/r}\alpha^h|A|.$$

Letting r tend to infinity finishes off the proof. □

A few remarks. Ruzsa in fact showed that there exists a nonempty subset $\emptyset \neq Y \subseteq A$ such that $|Y + B_1 + B_2| \leq \alpha^2|X|$.

It is worth comparing Ruzsa's bound with that given by Lemma 1.4. Setting $B = B_1$ and $C = B_2$ in the lemma yields a non-empty $\emptyset \neq X \subseteq A$ such that

$$|X + B_1 + B_2| \leq \frac{|X + B_1|}{|X|}|X + B_2| \leq \frac{|A + B_1|}{|A|}|X + B_2| \leq \alpha|X + B_2|.$$

No information is known about $|X+B_2|$, so we bound it by $|A+B_2| \leq \alpha|A|$. Putting everything together gives

$$|X + B_1 + B_2| \leq \alpha^2|A|.$$

While this is a bound than that given by Ruzsa, in practice the difference between having $|X|$ and $|A|$ on the right side is not important.

The careful reader will have noted that we in fact just gave a simpler proof of the lemma.

3 Covering lemmas

So far have seen that if $|A+B|$ is "small" compared to $|A|$, then we can say something about the cardinality of higher sum-and-difference sets.

Now we want to do a little more: starting from the condition that $|A+B|$ is "small", we want to cover B by "few translates of A". What does cover a set by translates of another set mean?

Definition. Let A and B be sets in a commutative group. B is *covered* by k translates of A if there exist elements $s_1, \dots, s_k$ in the ambient group such that

$$B \subseteq \bigcup_{i=1}^{k} (s_i + A) \text{ or equivalently } B \subseteq S + A, \text{ where } S = \{s_1, \dots, s_k\}.$$

Real time exercise. For each of the following you are given two sets A and B. Find a set S of least cardinality such that $B \subseteq S + A$.

(i) $A = \{1, \dots n\}$ and $B = \{1, \dots, n+1\}$.

(ii) $A = \{1, \dots n\} \times \{0\}$ and $B = \{1, \dots, n\} \times \{1, \dots, n\}$.

(iii) $A = \{1, \dots n\}$ and $B = \{n^2, 2n^2, \dots, n^3\}$.

Let's see what is going on. Set $\alpha = |A+B|/|A|$.

(i) $\alpha = 2$ and we needed 2 translates.

(ii) α is about $2n$ and we needed n translates.

(iii) α is about n and we needed n translates.

Your conjecture is:

Consider now one more example.

(iv) A is a random subset of $\{1, \ldots, n\}$ and $B = \{1, \ldots, n\}$.

What is a random subset of $\{1, \ldots, n\}$? For each $i = 1, \ldots, n$ flip a coin; if you get H put i in A and if you get T do not put i in A. Formally: each integer between 1 and n is included in A with uniform probability $1/2$ independent of all the others.

In this case α is about 4. This is because the expected value of $|A|$ is $n/2$. With a little work one can show that it is very likely that $|A|$ is about $n/2$. For all subsets A we have $A + B \subseteq \{2, \ldots, 2n\}$ and so $|A + B| \leq 2n$, which is about $4|A|$ for most random sets.

An exercise in probabilistic arguments shows that B cannot be covered by fewer than $\log(n)$ translates of A.

Note however, that $A - A$ is very likely to include $\{-\lceil n/3 \rceil, \ldots, \lceil n/3 \rceil\}$. There are at least $2n/3$ ways to express $i \in \{-\lceil n/3 \rceil, \ldots, \lceil n/3 \rceil\}$ as a difference $i = b - b'$ with $b, b' \in B$. So the expected number of representations of $i = a - a'$ where $a, a' \in A$ is at least $(1/2)^2 2n/3 = n/6$. When n is large, it is extremely likely that <u>all</u> such i lie in $A - A$.

So B can be covered by α translates of A.

Difference sets are nice sets – they have few holes

Lemma 3.1 (Ruzsa's covering lemma)**.** *Let $\alpha \in \mathbb{R}$ and A, B be finite non-empty sets in a commutative group. Suppose that $|A + B| \leq \alpha |A|$. Then $B \subseteq S + A - A$ where $S \subseteq B$ satisfies $|S| \leq \lfloor \alpha \rfloor$.*

Proof. The key is to select a maximal subset $S \subseteq B$ such that the sets $s + A$ are pairwise disjoint for all $s \in S$. It follows that

$$|S||A| = |A + S| \leq |A + B| \leq \alpha |A|.$$

The cardinality of S is a positive integer and so $|S| \leq \lfloor \alpha \rfloor$.

Let $b \in B$. There are two possibilities.

Either $b \in S$, in which case for any $a \in A$ we have $b = b + a - a \in S + A - A$.

Or $b \notin S$. By the maximality of S, this can only happen if $b + A$ intersects some $s + A$. In other words if there exists $s \in S$ and $a, a' \in A$ such that $b + a' = s + a$. This gives $b = S + a - a' \in S + A - A$, and completes the proof. □

Application 3.2 (Ruzsa)**.** *Let $\alpha \in \mathbb{R}$ and A be a finite non-empty set in a commutative group. Suppose that $|A+A| \leq \alpha|A|$. Then for all positive integers $h \geq 2$*

$$|hA - A| \leq \binom{\alpha^4 + h - 2}{h-1} \alpha^2 |A|.$$

Proof. This is a combination of Ruzsa's covering lemma and the Plünnecke–Ruzsa inequalities of Theorem 1.6.

To begin we note that the Plünnecke–Ruzsa inequalities give $|A+(2A-A)| = |3A-A| \leq \alpha^4|A|$.

Applying the covering lemma we get $2A - A \subseteq S + A - A$ for some $|S| \leq \alpha^4$.

An inductive argument gives $hA - A \subseteq (h-1)S + (A-A)$:

$$\begin{aligned} hA - A &= [(h-1)A - A] + A \\ &\subseteq [(h-2)S + (A-A)] + A \\ &= (h-2)S + (2A - A) \\ &\subseteq (h-2)S + S + A - A \\ &= (h-1)S + (A-A). \end{aligned}$$

Therefore

$$|hA - A| \leq |(h-1)S + (A-A)| \leq |(h-1)S|\,|A-A|.$$

The first term in the product is bounded above by $\binom{|S| + h - 2}{h-1} \leq \binom{\alpha^4 + h - 2}{h-1}$. The second term is at most $\alpha^2|A|$ by Lemma 1.2, as required. □

Later on in the course a covering lemma of Chang will be used, which is more efficient in applications.

Lemma 3.3 (Chang's covering lemma)**.** *Let $\alpha, \beta \in \mathbb{R}$ and A, B be finite non-empty sets in a commutative group. Suppose that $|A+B| \leq \beta|B|$ and $|A+A| \leq \alpha|A|$. Then there exists*

$$t \leq 1 + \lfloor \log_2(\alpha\beta)) \rfloor \leq 2\log(\alpha\beta)$$

and finite subsets $S_1, \ldots, S_t \subseteq A$ of cardinality at most $2\lfloor \alpha \rfloor$ such that

$$A \subseteq B - B + S_t + (S_{t-1} - S_{t-1}) + \cdots + (S_1 - S_1).$$

Remark. One can add 0 to S_t to get the more symmetric form

$$A \subseteq B - B + (S_t - S_t) + (S_{t-1} - S_{t-1}) + \cdots + (S_1 - S_1).$$

As we will see the proof gives $|S_t| < 2\lfloor\alpha\rfloor$ and so $|S_t \cup \{0\}| \leq 2\lfloor\alpha\rfloor$. Note, however, that it is no longer necessarliy the case that $A_t \cup \{0\} \subseteq A$.

Proof of Lemma 3.3. For simplicity we take α to be an integer.

We construct a sequence of sets $B = B_1, B_2, \ldots, B_t$ of increasing cardinality as follows.

Set $B_1 = B$. For $i > 1$ we check whether a subset $S_i \subseteq A$ of size 2α exists with the property that $|S_i + B_i| = |S_i||B_i|$.

If such S_i exists, we set $B_{i+1} = S_i + B_i$ and proceed.

If no such S_i exists, we select a maximal subset $S_i \subseteq A$ subject to $|S_i + B_i| = |S_i||B_i|$, set $B_{i+1} = S_i + B_i$ and terminate the algorithm.

Our first task is to prove that the algorithm terminates in at most $\lfloor(1 + \log_2(\alpha\beta))\rceil$ steps.

Suppose the algorithm continues at step i. Then

$$B_{i+1} = S_i + B_i = \cdots = S_i + \cdots + S_1 + B_1 = S_i + \cdots + S_1 + B. \tag{3.1}$$

The S_i have been chosen in such a way that

$$|B_{i+1}| = |S_i| \ldots |S_1||B| = (2\alpha)^i|B|.$$

Now observe that $S_j \subseteq A$ and so $B_{i+1} \subseteq iA + B$. The twin to Ruzsa's triangle inequality (Lemma 1.7) and Plünnecke's inequality (Theorem 1.5) gives

$$|iA + B| \leq \frac{|iA + A|}{|A|}|A + B| \leq \alpha^{i+1}\beta|B|.$$

Therefore

$$(2\alpha)^i|B| \leq \alpha^{i+1}\beta|B|$$

and so $i \leq \log_2(\alpha\beta)$. Therefore the algorithm must terminate after at most $1 + \log_2(\alpha\beta)$ steps.

At this ultimate step t, the maximality of S_t implies that for every $a \in A$ there exists $s \in S_t$ such that $(a + B_t) \cap (s + B_t) \neq \emptyset$. Therefore, $a \in S_t + B_t - B_t$ and consequently $A \subseteq S_t + B_t - B_t$.

Equation (3.1) implies

$$A \subseteq S_t + (S_{t-1} - S_{t-1}) + B_{t-1} - B_{t-1} = \cdots = S_t + (S_{t-1} - S_{t-1}) + \cdots + (S_1 - S_1) + B - B,$$

as claimed. □

What if we want to cover B by just translates of A? We succeed, but at a cost.

Lemma 3.4 (Covering by translates of the set)**.** *Let $\alpha \in \mathbb{R}$ and A, B be finite non-empty sets in a commutative group. Suppose that $|A - B| \leq \alpha|A|$. Then there exists a set $S \subseteq B - A$ of cardinality at most $\lceil \alpha \log(|B|) \rceil$ such that $B \subseteq S + A$.*

Remarks. The example with the random subset found at the top of p.11 suggests that the bound on the number of necessary translates is sharp. The $\log(|B|)$ factor can in general make a huge difference. However, if we take cardinalities, then the power trick allows us to drop this additional factor. So in some circumstances this last covering lemma turns out to be the most efficient.

There are other similar covering lemmas. For example if $|A + B| \leq \alpha|A|$ then one can find $S \subseteq B - A$ of cardinality $O(\alpha \log(|B|))$ such that $B \subseteq S + A$. The proofs of such results use additive energy and are not covered here.

The proof of the lemma is postponed until Section 5.

4 Freiman isomorphisms

Combinatorics is not enough to get a strong version of Freiman's theorem. One needs to be able to perform Fourier analysis on $\mathbb{Z}_n := \mathbb{Z}/n\mathbb{Z}$, the integers modulo n.

For technical reasons it is often much better to do Fourier analysis on $\mathbb{Z}_n$ rather than $\mathbb{Z}$. It is even more advantageous to do Fourier analysis on the characteristic function of a set whose relative density in $\mathbb{Z}_n$ is not too small.

With this in mind we set a goal: start with a finite set $A \subset \mathbb{Z}$ and produce a "model" B for A. B will be a "somewhat dense" subset of $\mathbb{Z}_n$ for some n and crucially "encode all the additive structure of A".

Have to wait a week or two to see why this is a sound strategy. At this stage we only introduce one important notion and prove one important result.

Let us begin with a definition due to Freiman, which captures in a concise way the phrase "B encodes all the additive structure of A".

Definitions. Let $k \geq 2$ be a positive integer and A a subset of a commutative group. A map $\phi : A \mapsto H$ from A to a commutative group H is a *k-Freiman homomorphism* if for all $x_1, \ldots, x_{2k} \in A$, the condition

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

implies that

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

ϕ is a *k-Freiman isomorphism* if it is an injection and the condition

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

is equivalent to

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

When no subscript appears, we assume $k = 2$.

We say A is *k-Freiman isomorphic* to $\phi(A)$.

Note. A Freiman homomorphism is a map with domain A.

A strange definition, so let us familiarise ourselves with it.

Real time exercise. For each of the following you are given k, A, H and ϕ. Decide whether ϕ is a freiman k-homomorphism. If it is, decide whether it is a k-Freiman isomorphism.

(i) Any k, any A, any H and ϕ the trivial map that maps everything to zero (in H).

(ii) Any k, $A = \{1, \dots, n\} \subset \mathbb{Z}$, $H = \mathbb{Z}$ and $\phi(i) = 2i$.

(iii) $k = 2$, $A = \{0, 1\} \subset \mathbb{Z}$, $H = \mathbb{Z}_2$ and ϕ is "reduction mod 2" $\phi(i) = i \bmod 2$.

(iv) Any k, any $A \subset \mathbb{R}^d$, $H = \mathbb{R}^d$ and $\phi(\mathbf{v}) = A\mathbf{v} + \mathbf{b}$ for some invertible $d \times d$ matrix A and some $\mathbf{b} \in \mathbb{R}^d$.

Solution.

Example 4.1. *(i) Let $k, n \geq 2$ be positive integers and $A \subset \mathbb{Z}$ a finite set of positive integers. Reduction* mod$n$*, $\phi : A \mapsto \mathbb{Z}_n$ given by $\phi(i) = i \bmod n$, is a k-Freiman homomorphism. It is a k-Freiman isomorphism if there is no wrap-around: $k \max\{A\} < n$.*

(ii) Let $k \geq 2$ be a positive integer, p be a prime, $0 \neq q \in \mathbb{Z}_p$ and $A \subseteq \mathbb{Z}_p$ a set of residues. Multiplication by q is a k-Freiman isomorphism, $\phi : A \mapsto \mathbb{Z}_p$ given by $\phi(i) = qi \bmod p$.

(iii) Let $n, k \geq 2$ be positive integers and $A \subseteq \mathbb{Z}_n$ a finite set of residues. Mapping $x \in \mathbb{Z}_n$ to the unique residue in $\{0, \ldots, n-1\}$ is a k-Freiman isomorphism provided that $A \subseteq I_j = (\frac{(j-1)n}{k}, \frac{jn}{k}]$. $\phi : A \mapsto \mathbb{Z}$ defined by $\phi(x) = x$.

Proof. In principle, we must establish two properties: ϕ is an injection and

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

In fact must only check the if and only if statement!

Injectivity of ϕ follows from $\Longleftarrow$ by taking $x_1 = \cdots = x_k = x$ and $x_{k+1} = \cdots = x_{2k} = y$: If $\phi(x) = \phi(y)$, then $k\phi(x) = k\phi(y)$ so $kx = ky$. When the ambient group is $\mathbb{Z}$ of $\mathbb{Z}_p$, "k is cancelled" and so we get $x = y$.

(i) Only show the second part. Suppose that $x_1 + \cdots + x_k = x_{k+1} + \cdots + x_{2k}$. Then, as both sides are integers in $\{1, \ldots, k \max\{A\}\}$, we get that the equation is equivalent to $x_1 + \cdots + x_k \equiv x_{k+1} + \cdots + x_{2k} \bmod n$. We are done.

(ii) Suppose $x_1 + \cdots + x_k \equiv x_{k+1} + \cdots + x_{2k} \bmod p$. q is invertible so this is equivalent to $q(x_1 + \cdots + x_k) \equiv q(x_{k+1} + \cdots + x_{2k}) \bmod p$. This last equations is equivalent to $qx_1 + \cdots + qx_k \equiv qx_{k+1} + \cdots + qx_{2k} \bmod p$.

(iii) Suppose $x_1 + \cdots + x_k \equiv x_{k+1} + \cdots + x_{2k} \bmod n$. Note that because all the residues x_i lie in I_j, we get that the sum $x_1 + \cdots + x_k$, viewed as an integer, lies in $((j-1)n, n]$. So the statements $x_1 + \cdots + x_k \equiv x_{k+1} + \cdots + x_{2k} \bmod n$ and $x_1 + \cdots + x_k = x_{k+1} + \cdots + x_{2k}$ are equivalent. □

Proposition 4.2 (Ruzsa)**.** *Let $A \subset \mathbb{Z}$ be a finite non-empty set of integers and $\alpha \in \mathbb{R}$. Suppose that $|A+A| \leq \alpha|A|$. Let $k \geq 2$ be a positive integer and $m > \alpha^{2k}|A|$ be another positive integer. There exists a subset $A' \subseteq A$ of cardinality at least $|A|/k$ that is k-Freiman isomorphic to a subset of $\mathbb{Z}_m$.*

Proof. We construct a series of Freiman isomorphisms ϕ_1, ϕ_2, ϕ_3 and ϕ_4 and take their composition $\mathbb{Z} \xrightarrow{\phi_1} \mathbb{Z}_p \xrightarrow{\phi_2} \mathbb{Z}_p \xrightarrow{\phi_3} \mathbb{Z} \xrightarrow{\phi_4} \mathbb{Z}_m$.

- ϕ_1 is reduction mod$p$. We select any $p > k \max\{A\}$ and so get a Freiman k-isomorphism.

- ϕ_2 is multiplication by $q \not\equiv 0 \bmod p$. A Freiman k-isomorphism. We need to select q carefully. We will show later on that by cardinality considerations, a suitable q will exist.

- ϕ_3 is the map $x \mapsto [x]_p$ where a residue $x \bmod p$ is mapped to its representative in $\{0, 1, \ldots, p-1\}$. This is a k-Freiman isomorphism provided that its domain is a subset of I_j for some $j \in \{1, \ldots, k\}$. We will select a suitable j that will depend on p and q later on. This choice will determine A'.

- ϕ_4 is reduction mod$m$. We will show that for all p and the particular q, j chosen, any m works provided that $m > \alpha^{2k}|A|$.

There are two outstanding issues. Let us clear up the one corresponding to ϕ_3. Let

$$A_j = \phi_1^{-1}\phi_2^{-1}(A \cap I_j) = \{a \in A : \phi_2\phi_1(a) \in I_j\}.$$

The A_j partition A and so $\sum_{j=1}^{k} |A_j| = |A|$. The average of the $|A_j|$ is therefore $|A|/k$. So there is a j, which depends on p and q, such that $|A_j| \geq |A|/k$. We set $A' = A_j$.

The ultimate issue is ensuring that

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

Let us start from the right side.

$$\begin{aligned}
&\phi(x_1) + \cdots + \phi(x_k) = \phi(x_{k+1}) + \cdots + \phi(x_{2k}) \iff \\
&[qx_1]_p + \cdots + [qx_k]_p \equiv [qx_{k+1}]_p + \cdots + [qx_{2k}]_p \mod m \iff \\
&[qx_1]_p + \cdots + [qx_k]_p - [qx_{k+1}]_p - \cdots - [qx_{2k}]_p \equiv 0 \mod m \iff \\
&[q(x_1 + \cdots + x_k - x_{k+1} - \cdots - x_{2k})]_p \equiv 0 \mod m \text{ (because all } x_i \in A_j).
\end{aligned}$$

So for ϕ to be a k-Freiman isomorphism we require that the only solution to the above equation is when $z := x_1 + \cdots + x_k - x_{k+1} - \cdots - x_{2k} = 0$.

Fix any $0 \neq z \in kA - \ell A$. As q ranges in $\{1, \ldots, p-1\}$, the product $[qz]_p$ also ranges in $\{1, \ldots, p-1\}$. So there are at most $(p-1)/m$ values of q where $[qz]_p$ is divided by m.

There are $|kA - \ell A| - 1 \leq (\alpha^{k+\ell}|A| - 1)$ such z and so at most

$$(\alpha^{k+\ell}|A| - 1)\frac{p-1}{m} < p - 1$$

values of q that will not work. So there is a value of q that works, provided that $m > \alpha^{k+\ell}|A|$. This finishes the proof. □

A generalisation of sorts of the previous result to sets in any commutative group was given b Green and Ruzsa.

Theorem 4.3 (Green-Ruzsa)**.** *Let A be a finite non-empty set in a commutative group and $\alpha \in \mathbb{R}$. Suppose that $|A+A| \leq \alpha|A|$. Then for all $k \geq 2$ there exists a group G of cardinality at most $C|A|$ such that A is k-Freiman isomorphic to a subset of G.*

C depends on k and α and may be taken to be $C = (10k\alpha)^{10\alpha^2}$.

5 Representation as sums and additive energy

The topic now becomes more tangible. Let A and B be finite non-empty sets in a commutative group. We study the number of representations of an element in the ambient group as a sum of elements in A and B.

It turns out there are a few equivalent ways to define this quantity.

$$\begin{aligned} r_{A+B}(x) &= \#\text{ representations of } x \text{ as a sum in } A+B \\ &= |\{(a,b) \in A \times B : x = a+b\}| \\ &= |(x-A) \cap B| \\ &= |A \cap (x-B)|. \end{aligned}$$

Note that $r_{A+B}(x) \leq \min\{|A|, |B|\}$.

r_{A+B} is *supported* on $A+B$ – the set of x where $r_{A+B}(x) \neq 0$ is $A+B$.

Real time exercise. (i) Let $A = B = \mathbb{Z}_p$. Find $r_{A+B}(x)$, for all x in the support of r_{A+A}.

(ii) Let $A = \mathbb{Z}_p \times \{\mathbf{0}\}$ and $B = \{0\} \times \{\mathbf{e}_1 \ldots, \mathbf{e}_d\}$ in $\mathbb{Z}_p \times \mathbb{Z}^d$. Find $r_{A+B}(x)$, for all x in the support of r_{A+B}.

Solution.

Let us now compute the sum of $r_{A+B}(x)$ over all x in the ambient group.

$$\sum_x r_{A+B}(x) = \sum_{x \in A+B} r_{A+B}(x) = |A|\,|B|. \tag{5.1}$$

The justification is that each pair $(a,b) \in A \times B$ contributes exactly once to the sum (for $x = a+b$). One can also argue as follows.

$$\sum_x r_{A+B}(x) = \sum_x |(x-A) \cap B| = \sum_x \sum_{a \in A} 1_B(a-x) = \sum_{a \in A} \sum_x 1_B(a-x) = \sum_{a \in A} |B| = |A|\,|B|.$$

We now prove Lemma 3.4 on p.14.

Proof of Lemma 3.4. We construct a nested sequence of subsets of B: $B = B_1 \supset B_2 \supset \cdots \supset B_t = \emptyset$. The process terminates when B_t is empty.

At stage i we find $s_i \in B_i - A \subseteq B - A$ such that

$$B \setminus B_i \subseteq \{s_1, \ldots, s_i\} + A.$$

Setting $i = t$ gives $\{s_1, \ldots, s_t\}$ such that $B = B \setminus \emptyset \subseteq \{s_1, \ldots, s_t\} + A$.

The first step is completed as follows.

We want to find an s_1 such that a large part of B is contained in $s_1 + A$. It is natural to look at $|(x+A) \cap B|$. Identity (5.1) implies that the average value of the cardinalities of the non-empty intersections $(x + A) \cap B$ is

$$\frac{\sum_x |(x + A) \cap B|}{|A - B|} = \frac{|A||B|}{|A - B|} \geq \frac{|A||B|}{\alpha|A|} = \frac{|B|}{\alpha}.$$

So there exists $s_1 \in B - A$ such that $|(s_1 + A) \cap B| \geq |B|/\alpha$.

We set $B_2 = B \setminus (s_1 + A)$. Note that $|B_2| \leq \left(1 - \frac{1}{\alpha}\right) |B_1| = \left(1 - \frac{1}{\alpha}\right) |B|$.

We iterate. At stage i we have $B_i \subset B$ and want to find s_i such that $|(s_i + A) \cap B_i|$ is large. Identity (5.1) implies that the average value of the cardinalities of the non-empty intersections $(x + A) \cap B_i$ is

$$\frac{\sum_x |(x + A) \cap B_i|}{|A - B_i|} = \frac{|A||B_i|}{|A - B_i|} \geq \frac{|A||B_i|}{|A - B|} \geq \frac{|A||B_i|}{\alpha|A|} = \frac{|B_i|}{\alpha}.$$

So there exists $s_i \in B_i - A$ such that $|(s_i + A) \cap B_i| \geq |B_i|/\alpha$.

We set $B_{i+1} = B_i \setminus (s_i + A)$. Note that

$$|B_{i+1}| \leq \left(1 - \frac{1}{\alpha}\right) |B_i| \leq \left(1 - \frac{1}{\alpha}\right)^i |B| < \exp\left(-\frac{1}{\alpha}\right)^i |B| = \exp\left(-\frac{i}{\alpha}\right) |B|.$$

We keep going this way. Only left to estimate after how many steps the process must stop.

When $t = \lceil \alpha \log(|B|) \rceil$, $B_{t+1} = \emptyset$ as $|B_{t+1}| < 1$. So the process terminates in at most $\lceil \alpha \log(|B|) \rceil$ steps. □

We now introduce a further way to quantify additive structure, which involves the representation function and is correlated with small doubling.

Definition. Let A and B be finite non-empty sets in a commutative group. Their *additive energy* is

$$E(A,B) = \sum_{x \in A+B} r_{A+B}(x)^2.$$

We next examine some of the basic properties of the additive energy.

Lemma 5.1 (Cauchy–Schwarz lower bound on additive energy)**.** *Let A and B be finite non-empty sets in a commutative group. Then*

$$E(A,B) \geq \frac{|A|^2|B|^2}{|A+B|}.$$

In particular, if $|A+B| \leq \alpha|A|$, then $E(A,B) \geq \dfrac{|A|B|^2}{\alpha}$.

Proof. The proof is a combination of identity (5.1) and the Cauchy–Schwarz inequality. Let us see first the particular instance of the inequality we will apply.

$$\begin{aligned}\sum_{x \in S} a_x &= \sum_{x \in S} a_x \cdot 1 \\ &\leq \left(\sum_{x \in S} a_x^2\right)^{1/2} \left(\sum_{x \in S} 1\right)^{1/2} \\ &= \left(\sum_{x \in S} a_x^2\right)^{1/2} |S|^{1/2}.\end{aligned}$$

In particular $\displaystyle\sum_{x \in S} a_x^2 \geq \frac{\left(\sum_{x \in S} a_x\right)^2}{|S|}$. Therefore

$$\begin{aligned}E(A,B) &= \sum_{x \in A+B} r_{A+B}(x)^2 \\ &\overset{C-S}{\geq} \frac{\left(\sum_{x \in A+B} r_{A+B}(x)\right)^2}{|A+B|} \\ &\overset{(5.1)}{=} \frac{|A|^2|B|^2}{|A+B|}.\end{aligned}$$

□

Our next task is to the express additive in equivalent formulations, which, depending on the context, can be more convenient to use than the definition.

Lemma 5.2 (Equivalent definitions of additive energy)**.** *Let A and B be finite non-empty sets in a commutative group. Then*

$$\begin{aligned} E(A,B) &= \sum_{x+y=z+w} 1_A(x)1_B(y)1_A(z)1_B(w) \\ &= \sum_{x-w=z-y} 1_A(x)1_B(y)1_A(z)1_B(w) \\ &= \sum_{x\in A-B} r_{A-B}(x)^2 \\ &= \sum_{a\in A,b\in B} |(a+B)\cap(A+b)| \\ &= \sum_{a\in A,b\in B} |(a-B)\cap(A-b)|. \end{aligned}$$

Proof. The proof of the first identity is useful in many contexts.

$$\begin{aligned} \sum_s r_{A+B}(s)^2 &= \sum_s \left(\sum_{x+y=s} 1_A(x)1_B(y)\right)^2 \\ &= \sum_s \left(\sum_{x+y=s} 1_A(x)1_B(y)\right)\left(\sum_{z+w=s} 1_A(z)1_B(w)\right) \\ &= \sum_s \sum_{x+y=s=z+w} 1_A(x)1_B(y)1_A(z)1_B(w) \\ &= \sum_{x+y=z+w} 1_A(x)1_B(y)1_A(z)1_B(w). \end{aligned}$$

The second identity follows as $x+y=z+w \iff x-w=z-y$. The third now follows from the above calculation (replacing B by $-B$).

The fourth follows from the first, and the fifth from the third.

$$\begin{aligned} \sum_{x+y=z+w} 1_A(x)1_B(y)1_A(z)1_B(w) &= \sum_{a\in A,b\in B} \sum_{a+y=z+b} 1_B(y)1_A(z) \\ &= \sum_{a\in A,b\in B} |(a+B)\cap(A+b)|. \end{aligned}$$

□

The last two identities combined with the Cauchy–Schwarz lower bound and averaging arguments allow one to prove variants of Lemma 3.4. On of the four possible statements is.

Lemma 5.3. *Let $\alpha\in\mathbb{R}$ and A,B be finite non-empty sets in a commutative group. Suppose that $|A+B|\le\alpha|A|$. There exists $S\subseteq B-A$, $|S|\le\lceil\alpha\log(|B|)\rceil$ such that $B\subseteq S+A$.*

6 The Balog–Szemerédi–Gowers theorem

This celebrated theorem is a converse of sorts to the statement 'small doubling implies large additive energy'.

We have seen that 'small doubling' means a doubling constant not far off from the absolute minimum that is 1. So 'large additive energy' must mean additive energy not far off the absolute maximum. What is this absolute maximum?

Lemma 6.1. *Let A and B be finite sets in a commutative group. Their additive energy is bounded by each of the quantities $|A|^2|B|$, $|A||B|^2$ and $|A|^{3/2}|B|^{3/2}$.*

In particular $E(A,A) \leq |A|^3$.

Proof. We look at the first expression in Lemma 5.2.

$$E(A,B) = \sum_{x+y=z+w} 1_A(x)1_B(y)1_A(z)1_B(w)$$

For each triplet $(x,y,z) \in A \times B \times A$ there is at most one $w \in B$ such that $w = x+y-z$. Therefore the sum is bounded above by $|A|^2|B|$. Similarly $|A||B|^2$ is also an upper bound.

Considering the product of (both sides of) the two upper bounds and taking a square root yields the symmetric upper bound $E(A,B) \leq |A|^{3/2}|B|^{3/2}$. □

Let us now check what can we say about some sets with large additive energy.

Real time exercise. For each of the following sets estimate: the additive energy, the doubling constant, the largest cardinality of a subset that has small doubling – here take small to mean a number smaller than any power of the cardinality of the subset.

(i) $A = \mathbb{Z}_n$.

(ii) $A = \mathbb{Z}_n \times \{0\} \cup \{0\} \times \mathbb{Z}_n \subset \mathbb{Z}_n^2$. A can be written in a convenient if sloppy way as $(\mathbb{Z}_n, 0) \cup (0, \mathbb{Z}_n)$.

(iii) $A = (\mathbb{Z}_n, 0, \ldots, 0) \cup (0, \mathbb{Z}_n, 0, \ldots, 0) \cup \cdots \cup (0, \ldots, 0, \mathbb{Z}_n) \subseteq \mathbb{Z}_n^d$.

Solution.

The conclusions of the last example and Lemma 6 is that we cannot hope to do much better than the following statement: if $E(A,A) \geq \delta |A|^3$, then A must contain a subset A' of relative density at lest δ and doubling at most δ^{-1}.

The Balog–Szemerédi–Gowers is a statement like the above, where δ is replaced by a power of itself. Balog and Szemerédi first proved the theorem, but with much weaker bounds.

We will be a little sloppy in the statement and proof of the theorem by not keeping track of constants. We write $P \ll Q$ if there exists a constant C such that $P \leq CQ$ and $P \gg Q$ if there exists a constant C such that $P \geq CQ$.

Theorem 6.2 (Balog–Szemerédi–Gowers)**.** *Let $\delta > 0$ be a real number and A be a finite set in a commutative group. Suppose that $E(A,A) \geq \delta |A|^3$.*

There exists a subset $A' \subseteq A$ of cardinality at least

$$|A'| \gg \delta^{10}|A|$$

such that

$$|A' - A'| \ll \delta^{-32}|A'|.$$

Remark. There are more efficient versions of this result. The important fact is that in the conclusion only powers of δ or δ^{-1} appear.

Idea of the proof:

The first part described above is accomplished by a random selection process.

Lemma 6.3 (Gowers)**.** *Let m and n be positive integers and $\varepsilon > 0$ a positive real number. Suppose $A_1, \dots, A_m$ are sets in $\{1, \dots, n\}$ such that $\sum_{i=1}^{m} |A_i| \geq \varepsilon nm$.*

There exists a subset $B \subseteq \{1, \dots, m\}$ of size at least $|B| \geq \frac{\varepsilon^5}{2} m$ such that for at least 90% of pairs $(i, j) \in B \times B$, $|A_i \cap A_j| \geq \frac{\varepsilon^2}{2} n$.

Proof. The set B is chosen by a random process. We let x_1, x_2, x_3, x_4 and x_5 be chosen uniformly at random from $\{1, \dots, n\}$. B is the set (or more formally, the random variable)

$$B = \{i \in \{1, \dots, m\} : x_1, x_2, x_3, x_4, x_5 \in A_i\}.$$

Let us calculate a lower bound on the cardinality of B using the linearity of expectation.

$$\begin{aligned}
\mathbb{E}[|B|] &= \sum_{i=1}^{m} \Pr(\{x_1, x_2, x_3, x_4, x_5\} \in A_i) \\
&= \sum_{i=1}^{m} \Pr(x_1 \in A_i)^5 \\
&= \sum_{i=1}^{m} \left(\frac{|A_i|}{n} \right)^5 \\
&\overset{\text{Jansen}}{\geq} m \left(\sum_{i=1}^{m} \frac{|A_i|}{nm} \right)^5 \\
&\geq \varepsilon^5 m.
\end{aligned}$$

Side note: a perhaps more intuitive way to interpret the first inequality is to say that the sum of the fifth powers is minimised when all summands are equal to their average.

We deduce a lower bound on the cardinality of $B \times B$.

$$\mathbb{E}[|B \times B|] = \mathbb{E}[|B|^2] \geq \mathbb{E}[|B|]^2 \geq \varepsilon^{10} m^2. \tag{6.1}$$

We used the fact that the variance of B is non-negative (or the Cauchy–Schwarz inequality).

The nest step is to bound from above the expected number of pairs $(i, j) \in B \times B$ where $|A_i \cap A_j| \leq \frac{\varepsilon^2}{2} n$. We define $C \subset B \times B$ to be the set of "bad" ordered pairs:

$$C = \{(i, j) \in B \times B : |A_i \cap A_j| \leq \frac{\varepsilon^2}{2} n\}.$$

We want to bound the expected value of the cardinality of $|C|$. Once again the linearity of expectation is handy.

$$\mathbb{E}[|C|] = \sum_{\substack{i,j=1 \\ |A_i \cap A_j| \leq \varepsilon^2 n/2}}^{m} \Pr((i,j) \in B \times B).$$

Note that if $|A_i \cap A_j| \leq \frac{\varepsilon^2}{2} n$, then

$$\begin{aligned}
\Pr((i,j) \in B \times B) &= \Pr(i \in B \text{ AND } j \in B) \\
&= \Pr(\{x_1, x_2, x_3, x_4, x_5\} \subseteq A_i \cap A_j) \\
&= \left(\frac{|A_i \cap A_j|}{n}\right)^5 \\
&\leq \frac{\varepsilon^{10}}{32}.
\end{aligned}$$

Therefore

$$\mathbb{E}[|C|] \leq \frac{\varepsilon^{10}}{32} m^2. \tag{6.2}$$

Combining inequalities (6.1) and (6.2) we get that

$$\mathbb{E}[|B \times B| - 16|C|] = \mathbb{E}[|B \times B|] - 16\mathbb{E}[|C|] \geq \frac{\varepsilon^{10}}{2} m^2.$$

It follows that there is a set B (or if you prefer an instance of the random variable) such that

$$|B| \geq \sqrt{\frac{\varepsilon^{10}}{2} m^2} \geq \frac{\varepsilon^5}{2} m$$

and

$$|B \times B| - 16|C| \geq 0.$$

In particular

$$\frac{|(B \times B) \setminus C|}{|B \times B|} = \frac{|B \times B| - |C|}{|B \times B|} \geq \frac{15}{16} \geq \frac{9}{10}.$$

The proof is completed. □

Remarks. The method is called dependent random choice. The number of random points x_i is determined by the desired degree of accuracy – 5 corresponds to $15/16 \geq 9/10$.

Before we move to the main body of the proof, let us isolate as another introductory lemma, a statement whose proof s typical and very useful.

Lemma 6.4. *Let n be a positive integer, $\delta > 0$ a real number and S a finite set. Suppose that $f : S \mapsto \{0, \ldots, n\}$ is a function that satisfies three properties:*

(i) $f(x) \leq n$ *for all* $x \in S$ *(this is implied by the co-domain of* f*).*

(ii) $\sum_{x \in S} f(x) = n^2$.

(iii) $\sum_{x \in S} f(x)^2 \geq \delta n^3$.

There exist at least $\delta n/2$ elements of S where $f(x)$ is at least $\delta n/2$.

Proof. Note that

$$\begin{aligned}
\delta n^3 &\leq \sum_{x \in S} f(x)^2 \\
&= \sum_{f(x) \geq \delta n/2} f(x)^2 + \sum_{f(x) < \delta n/2} f(x)^2 \\
&\leq \sum_{f(x) \geq \delta n/2} f(x)^2 + \frac{\delta n}{2} \sum_{f(x) < \delta n/2} f(x) \\
&\leq \sum_{f(x) \geq \delta n/2} f(x)^2 + \frac{\delta n}{2} \sum_{x} f(x) \\
&= \sum_{f(x) \geq \delta n/2} f(x)^2 + \frac{\delta n^3}{2}.
\end{aligned}$$

Therefore

$$\begin{aligned}
\frac{\delta}{2} n^3 &\leq \sum_{f(x) \geq \delta n/2} f(x)^2 \\
&\leq n^2 \sum_{f(x) \geq \delta n/2} 1 \\
&\leq n^2 |\{x \in S : f(x) \geq \delta n/2\}|.
\end{aligned}$$

So the the set of $x \in S$ where $f(x)$ is at least $\delta n/2$ has cardinality at least $\delta n/2$. □

Proof of Balog–Szemerédi–Gowers. The proof is completed in three steps.

1. Construction of two graphs with vertex sets subsets of A and edges determined by subsets of $A - A$; definition of A'.
2. Counting paths of length four that start and end in A' in the first graph.
3. Comparing a lower bound and an upper bound on the number of 8-tuples $(z_1, \ldots, z_8) \in A^8$ such that $z_1 - z_2 + z_3 - \cdots + z_7 - z_8 \in A' - A'$.

Reference list:

- popular difference d: $r_{A-A}(d) \geq \delta|A|/2$.
- G graph with vertex set A and ax an edge iff ax is a popular difference.
- $B \subseteq A$ where at least 90% of pairs (b, c) satisfy $|\Gamma_G(b) \cap \Gamma_G(c)| \geq \delta^4|A|/32$. $|B| \ll \alpha^{10}|A|$.
- H graph with vertex set B and bc an edge iff $|\Gamma_G(b) \cap \Gamma_G(c)| \geq \delta^4|A|/32$.
- $A' \subseteq B$ is determined by $a' \in A'$ if $|\Gamma_H(a')| \geq 4|B|/5$. $|A'| \geq |B|/2 \gg \alpha^{10}|A|$.

Step 1. We apply Lemma 6.4 to the set $S = A - A$ and the function r_{A-A}. There exist at least $\delta|A|/2$ so-called *popular differences* with at least $\delta|A|/2$ representations.

We form a graph G with vertex set A and edges determined by ax is an edge iff $a - x$ is a non-zero popular difference. Note that the definition is symmetric and antireflexive. We denote by $\Gamma_G(a)$ the set of neighbours of a in G.

Let us bound from bellow the sum of the cardinalities of the neighbourhoods in G.

$$\sum_{a \in A} |\Gamma_G(a)| = 2|E| = \sum_{d \text{ popular}} r_{A-A}(d) \geq \left(\frac{\delta}{2}|A|\right)\left(\frac{\delta}{2}|A|\right) = \frac{\delta^2}{4}|A|^2.$$

Lemma 6.3, applied to the sets $\Gamma(a)$ and $m = n = |A|$, guarantees the existence of a set $B \subseteq A$ such that

$$|B| \gg \delta^{10}|A| \tag{6.3}$$

and at least 90% of pairs $(b, c) \in B \times B$ satisfy $|\Gamma_G(b) \cap \Gamma_G(c)| \geq \delta^4|A|/32$.

We now define a new graph H with vertex set B and edges determined by bc is an edge iff $|\Gamma_G(b) \cap \Gamma_G(c)| \geq \delta^4|A|/32$.

Let us bound from bellow the sum of the cardinalities of the neighbourhoods in H.

$$\sum_{b\in B}|\Gamma_H(b)| \geq \frac{9}{10}|B|^2.$$

We are finally in position to define A'.

$$A' = \left\{b \in B : |\Gamma_H(b)| \geq \frac{4|B|}{5}\right\}.$$

A simple calculation shows that $|A'| \geq |B|/2$.

$$\begin{aligned}\frac{9}{10}|B|^2 &\leq \sum_{b\in B}|\Gamma_H(b)| \\ &= \sum_{b\in A'}|\Gamma_H(b)| + \sum_{b\notin A'}|\Gamma_H(b)| \\ &\leq |B||A'| + \frac{4|B|}{5}(|B| - |A'|).\end{aligned}$$

so that

$$|A'| \geq \frac{|B|}{2} \overset{(6.3)}{\gg} \delta^{10}|A|. \tag{6.4}$$

Step 2. Let us now bound from below the number of distinct paths of length four in the first graph G that start at a pair of distinct vertices $a' \neq b' \in A'$. Inclusion-exclusion gives that

$$|\Gamma_H(a') \cap \Gamma_H(b')| = |\Gamma_H(a') + |\Gamma_H(b')| - |\Gamma_H(a') \cup \Gamma_H(b')| \geq \frac{4|B|}{5} + \frac{4|B|}{5} - |B| = \frac{3|B|}{5}.$$

Therefore a' and b' have at least $3|B|/5$ common neighbours c.

By the definition of H we know that a' and c have at least $\delta^4|A|/32$ common neighbours x in G. Similarly b' and c have at least $\delta^4|A|/32$ common neighbours y in G.

Therefore there are at least

$$\left(\frac{3|B|}{5}\right)\left(\frac{\delta^4|A|}{32}\right)\left(\frac{\delta^4|A|}{32}\right) \overset{(6.3)}{\gg} \delta^{18}|A|^3 \tag{6.5}$$

distinct paths of the form $a'xcyb'$ (there are at least $3|B|/5$ choices for c and at least $\delta^4|A|/32$ choices for each of x and y).

Step 3. Let us now bound in two different ways the number N of 8-tuples $(z_1, \ldots, z_8) \in A^8$ such that $z_1 - z_2 + z_3 - \cdots + z_7 - z_8 \in A' - A'$.

It is clear that $N \leq |A|^8$.

To get a lower bound on N, we bound from bellow how many 8-tuples exist for each distinct $d = a' - b' \in A' - A'$. Note that each path $a'xcyb'$ in G corresponds to the identity

$$a' - b' = (a' - x) + (x - c) + (c - y) + (y - b').$$

Each edge corresponds to a popular difference, and so we can express, say, $a'-x$ in at least $\delta|A|/2$ ways as $z_1 - z_2 \in A - A$. Therefore each path corresponds to at least $(\delta|A|/2)^4$ distinct 8-tuples. Combining this with inequality (6.5) implies that there is $\gg \delta^{22}|A|^7$ 8-tuples $(z_1, \dots, z_8) \in A^8$ such that $z_1 - z_2 + z_3 - \cdots + z_7 - z_8 = a' - b'$.

Summing over all $d' \in A' - A'$ gives

$$N \gg \delta^{22}|A|^7|A' - A'|.$$

Comparing this with the upper bound of $|A|^8$ yields

$$|A' - A'| \ll \delta^{-22}|A| \overset{(6.4)}{\ll} \delta^{-32}|A'|.$$

□

7 The Szemerédi–Trotter theorem

In an abrupt change of topic we now move to combinatorial geometry.

We wish to derive sharp estimates on the number of incidences between a finite set of lines and a finite set of points in the plane $\mathbb{R}^2$. More on this later on.

It turns out that in the proof we will need the notion of a drawing of a graph and of the crossing number of a graph.

Definition. Let be G a finite graph. A *drawing* is a map that takes vertices to points of $\mathbb{R}^2$ and edges to curves (smooth functions from $[0,1] \mapsto \mathbb{R}^2$) which start and end at the images of their endpoint-vertices.

Example. Let $G = (V, E)$ with $V = \{1, 2, 3, 4, 5\}$ and $E = \{\{1, 2\}, \{2, 4\}, \{2, 5\}, \{3, 4\}, \{4, 5\}\}$. Give two genuinely different drawings.

Definition. Let be G a finite graph. The *crossing number* of G is the least number of crossings in any drawing of G on the plane (the number of points where a pair of edges intersect, excluding intersections at vertices).

Let us consider G with vertex set $V = \{1, 2, 3, 4\}$ and edge set $E = \{\{1, 2\}, \{3, 4\}\}$.

$\mathrm{cr}(G) =$.

Real time exercise. For each of the following cycles determine the crossing number.

(i) C_3 with vertex set $V = \{1, 2, 3\}$ and edge set $E = \{\{1, 2\}, \{2, 3\}, \{3, 1\}\}$.

(ii) C_4 with vertex set $V = \{1, 2, 3, 4\}$ and edge set $E = \{\{1, 2\}, \{2, 3\}, \{3, 4\}, \{4, 1\}\}$.

Let us now quickly cover some facts about the so-called planar graphs.

Definition. A graph is called *planar* if its crossing number is zero.

For example, the 5-cycle C_5 is a planar graph, yet the complete graph on 5 vertices K_5 is not.

When a graph is planar, one can talk about its faces.

A *face* of a finite planar graph is a connected region of the subset of the plane resulting from removing a drawing of the graph.

In the examples we saw above the faces are:

A word of warning: an intuitive interpretation of the word 'face' result in one fewer face than the definition. The definition forces a triangle to have two faces. All finite planar graphs have an unbounded face, which comes from the unbounded component.

Euler proved that for finite connected planar graphs.

$$|F| - |E| + |V| = 2. \tag{7.1}$$

Here $|F|$ is the number of faces (and as usual $|E|$ the number of edges and $|V|$ the number of vertices).

Let us deduce an upper bound on the number of edges in a finite planar graph, which will be useful later on. The controversial faces will not appear.

Lemma 7.1. *Let $G = (V, E)$ be a finite planar graph. Then $|E| < 3|V|$.*

Proof. We may assume that the graph is connected by considering the connected parts of the graph and adding both parts of the corresponding inequalities.

We show that $3|F| \leq 2|E|$. Let us count in two ways the number N of edge-face incidences. Each face is incident to at least three edges, while each edge is incident to two faces. Therefore

$$3|F| \leq N = 2|E|.$$

Euler's formula (7.1) now implies

$$0 < |F| - |E| + |V| \leq 2|E|/3 - |E| + |V| = -|E|/3 + |V|.$$

□

Corollary 7.2. *Let $G = (V, E)$ be a finite graph. Then $|E| < 3|V| + \mathrm{cr}(G)$.*

Proof. Suppose that we have a drawing of the finite graph with exactly $\mathrm{cr}(G)$ crossings. By removing at most one edge from each crossing we make the graph planar. Applying Lemma 7.1 to the resulting graph, which has $|V|$ vertices and at least $|E| - \mathrm{cr}(G)$ edges, gives

$$|E| - \mathrm{cr}(G) < 3|V|.$$

□

We now prove a lower bound on the number of crossings.

Lemma 7.3 (Ajtai–Chvátal–Newborn–Szemerédi, Leighton)**.** *Let $G = (V, E)$ be a finite graph. Suppose $|E| \geq 4|V|$. Then*

$$\mathrm{cr}(G) \geq \frac{|E|^3}{64|V|^2}.$$

Proof. The key is to consider induced random subgraphs of G. Let $G_p = (V_p, E_p)$ be the random induced subgraph where each vertex is included independently with uniform probability p.

Corollary 7.2 gives that the random variable $3|V_p| + \mathrm{cr}(G_p) - |E_p|$ is positive. Taking expectation and using its linearity gives

$$\mathbb{E}[|E_p|] - 3\mathbb{E}[|V_p|] < \mathbb{E}[\mathrm{cr}(G_p)].$$

Another application of the linearity of expectation gives that the left side is $p^2|E| - 3p|V|$.

The right side is at most $p^4\mathrm{cr}$. Each crossing in a drawing of G that gives rise to $\mathrm{cr}(G)$ appears with probability p^4. Therefore the expected number of crossings in the drawing of G_p coming from the aforementioned drawing of G is $p^4\mathrm{cr}(G)$. The expected value of $\mathrm{cr}(G_p)$ is bounded above by this quantity.

We therefore have

$$\mathrm{cr(G)} > \frac{p|E| - 3|V|}{p^3}.$$

Letting $p = 4|V|/|E|$, which by the hypothesis is at most one, gives the desired lower bound. □

Remark. The bound is attained up to a constant on K_n, the complete graph on n vertices.

We use this inequality to establish an upper bound on the number of point-line incidences.

Definition. Let P be a finite set of points and L be a finite set of lines on the plane $\mathbb{R}^2$. The number of point-line incidences is

$$I(P, L) = |\{(p, \ell) : p \in P, \ell \in L, p \in \ell\}|.$$

Examples.

Theorem 7.4 (Szemerédi–Trotter)**.** *Let P be a finite set of points and L be a finite set of lines on the plane $\mathbb{R}^2$. Then number of point-line incidences is at most*

$$I(P, L) \leq 4(|L|^{2/3}|P|^{2/3} + |L| + |P|).$$

Proof by Székely. Given a finite set of points and a finite set lines, we deduce a drawing of a graph by placing an edge between consecutive points on a line. This then gives rise to an abstract graph $G = (V, E)$. Note that $V = P$.

If $|E| \geq 4|V|$, then Lemma 7.3 gives

$$\frac{|E|^3}{64|V|^2} \leq \mathrm{cr}(G) \leq \binom{|L|}{2} \leq |L|^2.$$

The upper bound on the crossing number comes from the fact that any two distinct lines on the plane meet in at most one point. So there cannot be more crossings in G than the number of pairs of lines.

This yields $|E| \leq 4|L|^{2/3}|V|^{2/3} = 4|L|^{2/3}|P|^{2/3}$.

The bound

$$|E| \leq 4(|L|^{2/3}|P|^{2/3} + |P|).$$

accounts for the possibility that $|E| \leq 4|V|$.

The last task is to obtain an expression for $|E|$. Each line with i points incident on it, contributes precisely $(i - 1)$ edges to G and so $|E| = I(P, L) - |L|$. The claim now follows. □

Remark. All three terms are necessary.

The Szemrédi–Trotter theorem has plenty of applications in unexpected places. The most spectacular has to be that of Elekes on the sum-product problem of Erdős.

Conjecture 7.5. *Let $A \subset \mathbb{R}$ be a finite set of real numbers. Then*

$$\max\{|A \cdot A|, |A + A|\} \gg |A|^{2-\varepsilon} \text{ for all } \varepsilon > 0.$$

Remark. The ε is necessary. For example when $A = \{1, \ldots, n\}$, then $|A \cdot A| \ll |A|^2/\log\log(n)$. This is a non-trivial statement. It can be deduced by applying Chebyshev's inequality to the function $\omega(i)$ that counts the number of distinct prime factors of the positive integer i. A result of Erdős and Kac states that $\omega(n)$ (viewed as a random variable under the uniform distribution on $\{1, \ldots, n\}$) has asymptotic mean $\log\log(n)$ and asymptotic variance $\sigma^2 = \log\log(n)$.

Theorem 7.6 (Elekes)**.** *Let $A \subset \mathbb{R}$ be a finite set of real numbers. Then*

$$|A \cdot A|\,|A + A| \geq \frac{|A|^{5/2}}{16}.$$

In particular

$$\max\{|A \cdot A|, |A + A|\} \geq \frac{|A|^{5/4}}{4}.$$

Proof. We apply Szemerédi–Trotter to

$$P = (A + A) \times (A \cdot A) \text{ and } L = \{y = a(x - b) : a, b \in A\}.$$

Note that $|P| = |A + A||A \cdot A|$ and $|L| = |A|^2$. The line $y = a(x - b)$ is incident to the points $(b + c, ac)$ for all $c \in A$. Therefore each line is incident to at least $|A|$ points and so $I(P, L) \geq |A||L| = |A|^3$.

Szemerédi–Trotter gives

$$|A|^3 \leq 4(|A|^{4/3}|A + A|^{2/3}|A \cdot A|^{2/3} + |A|^2 + |A + A||A \cdot A|) \leq 5|A|^{4/3}|A + A|^{2/3}|A \cdot A|^{2/3}.$$

Consequently

$$|A + A||A \cdot A| \geq \frac{|A|^{5/2}}{5^{3/2}} \geq \frac{|A|^{5/2}}{16}.$$

□