SCHOEN'S PROOF OF THE BALOG-SZEMERÉDI-GOWERS THEOREM

1. Introduction

Let G be an abelian group. Writing 1_X to be the indicator function of the set X, we define, for every finite, non-empty $X, Y \subseteq G$, the convolution function $1_X \circ 1_Y : G \to \mathbb{R}$ as

$$1_X \circ 1_Y(n) = \sum_{a \in G} 1_X(n+a) 1_Y(a) \quad \text{for all} \quad n \in G.$$

We now define the additive energy $E(A)$ of a finite, non-empty subset A of G to be

$$E(A) = \sum_{n \in G} 1_A \circ 1_A(n)^2.$$

In particular, $E(A)$ counts the number of solutions to the equation $a_1 - a_2 = a_3 - a_4$, with $a_1, \dots, a_4 \in A$. We also define the difference set $A - A = \{a_1 - a_2 \mid a_1, a_2 \in A\}$. Since the function $1_A \circ 1_A$ is supported on $A - A$, we may apply Cauchy-Schwarz inequality on the definition of $E(A)$ to deduce that

$$E(A)|A - A| \geq (\sum_n 1_A \circ 1_A(n))^2 = |A|^4.$$

Thus, whenever $|A - A| \leq K|A|$, for some $K \geq 1$, then $E(A) \geq |A|^3/K$. It is natural to ask whether a converse holds true, that is, if $E(A) \geq |A|^3/K$, then is $|A - A| \leq K^C|A|$, for some absolute constant $C > 0$. This is false, since the set $A_N = \{1, \dots, N\} \cup \{2, 4, \dots, 2^N\}$ satisfies $E(A_N) \gg N^3 \gg |A_N|^3$ as well as $|A_N - A_N| \gg N^2 \gg |A_N|^2$.

The Balog-Szemerédi-Gowers theorem essentially tells us that these are the only counterexamples to the aforementioned heuristic.

Theorem 1.1. *Let $A \subseteq G$ be a finite set having $E(A) = |A|^3/K$, for some $K \geq 1$. Then there exists $A_1 \subseteq A$ such that*

$$|A_1| \gg |A|/K \quad \textit{and} \quad |A_1 - A_1| \ll K^3|A| \ll K^4|A_1|.$$

This is the version of Balog-Szemerédi-Gowers theorem which was proven by Schoen [1] and we present his proof in this note. We remark that one may similarly prove that

$$|A + A|E(A) \geq |A|^4,$$

whence, whenever $|A + A| \leq K|A|$, then $E(A) \geq |A|^3/K$. In fact, applying the Plünnecke–Ruzsa theorem to the conclusion of Theorem 1.1, one can show that whenever $E(A) \geq |A|^3/K$, then there exists $A_1 \subseteq A$ such that $|A_1| \gg |A|/K$ and $|A_1 + A_1| \ll K^8|A_1|$.

Furthermore, asymmetric versions of Theorem 1.1 are also known to hold. In particular, if we have two finite subsets A, B of some abelian group such that $|A| = |B| = n$ and

$$|\{(a_1, b_1, a_2, b_2) \in A \times B \times A \times B \ : \ a_1 + b_1 = a_2 + b_2\}| \geq n^3/K,$$

then there exist $A_1 \subseteq A$ and $B_1 \subseteq B$ such that $|A_1|, |B_1| \gg n/K^C$ and

$$|A_1 + B_1| \ll K^C n,$$

for some absolute constant $C > 0$. We do not pursue this here, but a proof of such a result is present in Chapter 6 of the lecture notes.

2. Proof of Theorem 1.1

The proof follows through three steps, we first reduce the main theorem to a lemma about showing that a significant part of A has most of its differences being "popular differences". In order to prove this lemma, we begin with a combinatorial set up and we then finish with some probabilistic ideas. We now present the first part, and thus, we record some preliminary definitions. For every $\tau \geq 1$, let

$$P_\tau = \{x \in G \mid 1_A \circ 1_A(x) \geq \tau\} \text{ and } Q_\tau = \{x \in G \mid \tau \leq 1_A \circ 1_A(x) < 2\tau\}.$$

Our goal in the first part is to prove Theorem 1.1 conditionally on the following lemma.

Lemma 2.1. *Let $c > 0$, let $A \subseteq G$ be a finite set with $E(A) = |A|^3/K$ Then there exists $A' \subseteq A$ with*

$$|A'| \geq |A|/6K \quad \textit{and} \quad \sum_{\substack{a,b \in A', \\ a-b \notin P_{c|A|/K}}} 1 \leq 16c|A'|^2.$$

Proof of Theorem 1.1. We apply Theorem 2.1 with $c = 1/128$ to obtain $A' \subseteq A$ such that $|A'| \geq |A|/6K$ and such that the set

$$S = \{(a, b) \in A' \times A' \mid a - b \in P_{c|A|/K}\}$$

satisfies $|S| \geq 7|A'|^2/8$. We now set A_1 to be all the elements $a \in A'$ such that there are at least $3|A'|/4$ choices of b with $(a, b) \in S$. In particular, this means that

$$7|A'|^2/8 \leq |S| = \sum_{a \in A_1} \sum_{b \in A} \mathbb{1}_{(a,b) \in S} + \sum_{a \in A' \setminus A_1} \sum_{b \in A} \mathbb{1}_{(a,b) \in S} \leq |A_1||A'| + (|A'| - |A_1|)(3|A'|/4).$$

Simplifying this gives us $|A_1| \geq |A'|/2 \geq |A|/12K$.

It now suffices to show that $|A_1 - A_1| \ll K^3|A|$. In order to show this, we first note that for any $a, b \in A_1$, there exist at least $|A'|/2$ choices of $y \in A'$ such that $(a, y), (b, y) \in S$. In order to see this, note that there are at most $|A'|/4$ "bad" choices of y for a and at most $|A'|/4$ "bad" choices of y for b, whence there are at most $|A'|/2$ choices of y which are "bad" for either of a or b. With this in hand, we see that for any $a - b \in A_1 - A_1$, there are at least $|A'|/2$ ways to write $a - b = (a - y) - (b - y)$. Moreover, since $(a, y) \in S$, there are at least $c|A|/K$ many ways to write $a - y$ as $a_1 - a_2$ with $a_1, a_2 \in A$. Noting a similar phenomenon for $b - y$, we see that

$$\sum_{x \in A_1 - A_1} |\{(a_1, \dots, a_4) \in A^4 \mid x = a_1 - a_2 - a_3 + a_4\}| \geq |A_1 - A_1|(|A'|/2)(c|A|/K)^2.$$

On the other hand, the left hand side here is bounded above by $|A|^4$, whereupon, we have

$$|A_1 - A_1| \leq 2c^{-2}K^2|A|^2/|A'| \leq 12c^{-2}K^3|A| \ll K^4|A_1|,$$

which is the desired bound. $\square$

Our aim for the rest of this section is to prove Lemma 2.1. We begin by performing some combinatorial pruning, and so, we note that

$$\sum_{x\notin P_{|A|/2K}} (1_A \circ 1_A)(x)^2 < (|A|/2K) \sum_x (1_A \circ 1_A)(x) \leq |A|^3/2K,$$

whence,

$$\sum_{x\in P_{|A|/2K}} (1_A \circ 1_A)(x)^2 = E(A) - \sum_{x\notin P_{|A|/2K}} (1_A \circ 1_A)(x)^2 > |A|^3/2K.$$

Writing $\mathcal{J} = \{j \in \mathbb{N} \ : \ |A|/4K \leq 2^j \leq |A|\}$, we see that

$$\sum_{j\in\mathcal{J}} |Q_{2^j}| 2^{2j+2} \geq \sum_{x\in P_{|A|/2K}} (1_A \circ 1_A)(x)^2 > |A|^3/2K. \tag{2.1}$$

Furthermore, we have that

$$\begin{aligned}\sum_{j\in\mathcal{J}} \sum_{\substack{a,b\in A,\\ a-b\notin P_{c|A|/K}}} |(a-A)\cap(b-A)\cap Q_{2^j}| &\leq \sum_{\substack{a,b\in A,\\ a-b\notin P_{c|A|/K}}} |(a-A)\cap(b-A)| \\ &= \sum_{\substack{a,b\in A,\\ a-b\notin P_{c|A|/K}}} (1_A \circ 1_A)(a-b) \\ &\leq c|A|^3/K \leq \sum_{j\in\mathcal{J}} 8c|Q_{2^j}|2^{2j},\end{aligned}$$

where the last inequality follows fom (2.1). This means that there exists $j \in \mathcal{J}$ such that

$$\sum_{\substack{a,b\in A,\\ a-b\notin P_{c|A|/K}}} |(a-A)\cap(b-A)\cap Q_{2^j}| \leq 8c|Q_{2^j}|2^{2j}. \tag{2.2}$$

We are now finished with our combinatorial set-up, and so, we proceed to the final part of the proof, where we employ the probabilistic method.

We fix some $j \in \mathcal{J}$ satisfying (2.2) and write $Q := Q_{2^j}$. We pick $s \in Q$ uniformly at random. In particular, this means that for any $X \subseteq G$, we have

$$\mathbb{P}(s \in X) = |X \cap Q|/|Q|.$$

Writing $A' = A \cap (A+s)$, we see that $a \in A'$ if and only if $a \in A$ and $a \in A+s$, where the latter condition is equivalent to $s \in a - A$. Thus, we have

$$\mathbb{P}(a \in A') = 1_A(a) \ |Q \cap (a-A)|/|Q|,$$

and so,

$$\mathbb{E}|A'| = |Q|^{-1} \sum_{a\in A} |Q \cap (a-A)| = |Q|^{-1} \sum_{q\in Q} (1_A \circ 1_A)(q) \geq 2^j. \tag{2.3}$$

Similarly,

$$\mathbb{P}(a, b \in A') = 1_A(a)1_A(b) \ |Q \cap (a-A) \cap (b-A)|/|Q|,$$

whence, writing $\mathcal{B} = \{(a,b) \in A' \times A' \mid a-b \notin P_{c|A|/K}\}$, we have

$$\mathbb{E}|\mathcal{B}| = |Q|^{-1} \sum_{\substack{a,b\in A,\\ a-b\notin P_{c|A|/K}}} |Q \cap (a-A) \cap (b-A)| \leq 8c2^{2j},$$

where the last inequality follows from (2.2). Combining this with (2.3) and the fact that $\mathbb{E}|A'|^2 \geq (\mathbb{E}|A'|)^2$, we get

$$\mathbb{E}(|A'|^2 - |\mathcal{B}|/16c) \geq 2^{2j}/2.$$

This means that there exists some $s \in Q$ such that

$$|A'|^2 - |\mathcal{B}|/16c \geq 2^{2j}/2,$$

which, in turn, gives us

$$|\mathcal{B}| \leq 16c|A'|^2 \quad \text{and} \quad |A'| \geq 2^j/2^{1/2} \geq |A|/(2^{5/2}K) \geq |A|/6K.$$

This finishes the proof of Lemma 2.1.

References

[1] T. Schoen, *New bounds in Balog-Szemerédi-Gowers theorem*, Combinatorica **35** (2015), no. 6, 695-701.