Chapter 3: The Balog-Szemerédi-Gowers Theorem

Adam Sheffer

April 12, 2016

1 Sets with a large energy

In this chapter we return to working over a general abelian group $\mathcal{G}$.

The Balog-Szemerédi-Gowers theorem is a main result of additive combinatorics, and has many equivalent formulations. In this section we present one such formulation and prove it. Then, in the following section we discuss additional formulations of the theorem.

In the previous chapter we introduced the notion of multiplicative energy of a set. We now symmetrically define the *additive energy* of a set $A \subset \mathcal{G}$:

$$E^+(A) = \left|\left\{(a_1,a_2,a_3,a_4) \in A^4 :\ a_1 + a_2 = a_3 + a_4\right\}\right|.$$

As with multiplicative energy, we have $|A|^2 \le E^+(A)$ since this is the number of solutions with $a_1 = a_3$ and $a_2 = a_4$. Similarly, $E^+(A) \le |A|^3$ since for any choice of a_1, a_2, and a_3 there is at most one valid choice for a_4.

For any $x \in A + A$ we set $r_A^+(x) = |\{(a_1,a_2) \in A^2 :\ a_1 + a_2 = x\}|$ and $r_A^-(x) = |\{(a_1,a_2) \in A^2 :\ a_1 - a_2 = x\}|$. Since every pair of A^2 participates in exactly one $r_A^+(x)$, we have $\sum_{x\in A+A} r_A^+(x) = |A|^2$. By the Cauchy–Schwarz inequality, we have

$$E^+(A) = \sum_{x \in A+A} r_A^+(x)^2 \ge \frac{\left(\sum_x r_A^+(x)\right)^2}{|A+A|} = \frac{|A|^4}{|A+A|}. \tag{1}$$

As with multiplicative energy, intuitively, the smaller $|A+A|$ is the larger $E^+(A)$ is. For example, an arithmetic progression has a large additive energy and a small difference set. A random set has a large sum set and a small additive energy.

By (1), a small value of $|A + A|$ always implies a large value of $E^+(A)$. However, the other direction does not always hold. For example, take $A = P \cup R$ where P is an arithmetic progression of size $|A|/2$ and R is a random set of size $|A|/2$. In this case we have $|A+A| = \Theta(|A|^2)$ and $E^+(A) = \Theta(|A|^3)$.

Although a large additive energy says nothing about the size of the sum set, it does provide a lot of information about the structure of the set. The following variant of the Balog-Szemerédi-Gowers theorem is by Schoen [1].

Theorem 1.1. *Let $A \subset \mathcal{G}$ such that $E^+(A) = \delta|A|^3$. Then there exists $A' \subset A$ such that $|A'| = \Omega(\delta|A|)$ and*

$$|A' - A'| = O\left(\delta^{-4}|A'|\right).$$

In other words, if a set A has a large additive energy then it must contain a large subset that has small doubling. Notice that we can obtain the same result for the case of a large multiplicative energy by applying Theorem 1.1 to $\lg A = \{\lg a : a \in A\}$. To prove Theorem 1.1, we first prove the following lemma.

For any $0 < \gamma < 1$, we set

$$A_\gamma(x) = \begin{cases} 1, & r_A^-(x) \geq \gamma|A|, \\ 0, & \text{otherwise}. \end{cases}$$

Lemma 1.2. *Let $A \subset G$ be a finite set and let $\delta > 0$ such that $E^+(A) = \delta|A|^3$. Then for any $c > 0$ there exists a set $A' \subset A$ with $|A'| \geq \delta|A|/3$ and*

$$\sum_{x \in A'-A'} r_{A'}^-(x) \cdot A_{c\delta}(x) \geq (1-16c)|A'|^2. \tag{2}$$

Intuitively, Lemma 1.2 states the following. If A has a large additive energy, then we can find a large subset $A' \subset A$ such that for many pairs $(a,b) \in A' \times A'$ the difference $a - b$ has many representations.

Proof of Lemma 1.2. The assertion of the lemma is trivial when $c \geq 1/16$. We may thus assume that $c < 1/16$. We will set $A' = A \cap (A+s)$ for some $s \in A - A$. First we perform a pruning step, removing bad candidates for s.

Since $\sum_{x \in A-A} r_A^-(x) = |A|^2$, we have

$$\sum_{\substack{x \in A-A \\ r_A^-(x) \leq \delta|A|/2}} r_A^-(x)^2 \leq \frac{\delta|A|}{2} \sum_{\substack{x \in A-A \\ r_A^-(x) \leq \delta|A|/2}} r_A^-(x) \leq \frac{\delta|A|^3}{2} = E^+(A)/2.$$

For an integer j, we set $Q_j = \{x \in A - A : |A|/2^{j+1} < r^-_A(x) \le |A|/2^j\}$. Since $\sum_{x\in A-A} r^-_A(x)^2 = E^+(A)$, we get

$$\sum_{j=0}^{\lg(2/\delta)-1} |Q_j| \left(\frac{|A|}{2^j}\right)^2 \ge \sum_{\substack{x\in A-A \\ r^-_A(x)>\delta|A|/2}} r^-_A(x)^2 \ge E^+(A)/2 = \frac{\delta}{2}|A|^3. \tag{3}$$

We also have

$$\sum_{j=0}^{\lg(2/\delta)-1} \sum_{\substack{(a,b)\in A^2 \\ A_{c\delta}(a-b)=0}} |(A-a)\cap(A-b)\cap Q_j| \le \sum_{\substack{(a,b)\in A^2 \\ A_{c\delta}(a-b)=0}} |(A-a)\cap(A-b)|$$

$$< |A|^2 \cdot c\delta|A| = c\delta|A|^3. \tag{4}$$

By combining (3) and (4) we obtain

$$\sum_{j=0}^{\lg(2/\delta)-1} |Q_j| \left(\frac{|A|}{2^j}\right)^2 > \frac{1}{2c} \sum_{j=0}^{\lg(2/\delta)-1} \sum_{\substack{(a,b)\in A^2 \\ A_{c\delta}(a-b)=0}} |(A-a)\cap(A-b)\cap Q_j|.$$

Thus, there exists $0 \le j \le \lg(2/\delta) - 1$ that satisfies

$$|Q_j| \left(\frac{|A|}{2^j}\right)^2 > \frac{1}{2c} \sum_{\substack{(a,b)\in A^2 \\ A_{c\delta}(a-b)=0}} |(A-a)\cap(A-b)\cap Q_j|,$$

or

$$\sum_{\substack{(a,b)\in A^2 \\ A_{c\delta}(a-b)=0}} |(A-a)\cap(A-b)\cap Q_j| < \frac{c|Q_j||A|^2}{2^{2j-1}}. \tag{5}$$

For simplicity, we write $Q = Q_j$. We uniformly choose $s \in Q$. We denote by $1_A : \mathcal{G} \to \{0,1\}$ the *indicator function* of A. That is, $1_A(x) = 1$ if $x \in A$ and otherwise $1_A(x) = 0$. For any $x \in \mathcal{G}$, we have

$$\Pr[s = x] = \frac{1_Q(x)}{|Q|}.$$

Similarly to $r^+_A(x)$, we set $r^+_{A,B}(x) = |\{(a,b) \in A \times B : a + b = x\}|$. Set $A' = A \cap (A + s)$. For every $a \in \mathcal{G}$ we have that $a \in A'$ if and only if $a \in A$ and $s \in a - A$. Thus,

$$\Pr[a \in A'] = 1_A(a)\frac{|(a-A)\cap Q|}{|Q|} = 1_A(a)\frac{r^+_{A,Q}(a)}{|Q|}.$$

By linearity of expectation, the above implies

$$E[|A'|] = \sum_{a \in A} \frac{r^+_{A,Q}(a)}{|Q|} = \sum_{q \in Q} \frac{r^-_A(q)}{|Q|} > \frac{|A|}{2^{j+1}}.$$

For any two elements $a, b \in \mathcal{G}$ we have that $a, b \in A'$ if $a, b \in A$ and $s \in (a - A) \cap (b - A)$. That is, for any $a, b \in A$ we have

$$\Pr\left[(a,b) \in (A')^2\right] = \frac{|(a - A) \cap (b - A) \cap Q|}{|Q|}.$$

We set $B = \{(a, b) \in (A')^2 : A_{c\delta}(a - b) = 0\}$. By combining the above with (5), we get

$$E[|B|] = \sum_{\substack{(a,b) \in A^2 \\ A_{c\delta}(a-b)=0}} \Pr\left[(a,b) \in (A')^2\right] = \sum_{\substack{(a,b) \in A^2 \\ A_{c\delta}(a-b)=0}} \frac{|(a - A) \cap (b - A) \cap Q|}{|Q|} \leq \frac{c|A|^2}{2^{2j-1}}.$$

We now see the intuition behind the choice of $Q = Q_j$ above — we chose a subset of $A - A$ such that all of the elements in it have about the same number of representations, and that the resulting set B is expected not to be very large. Since $E[|A'|^2] \geq E[|A'|]^2$, we obtain

$$E\left[|A'|^2 - |B|/16c\right] > \left(\frac{|A|}{2^{j+1}}\right)^2 - \frac{c|A|^2}{2^{2j-1}16c} = \frac{|A|^2}{2^{2j+3}}.$$

By this expectation, there exists a choice of s for which $|A'|^2 - |B|/16c \geq \frac{|A|^2}{2^{2j+3}}$. Since $j \leq \lg(2/\delta) - 1$, we get $|A'| \geq |A|/2^{j+3/2} > \delta|A|/3$ and $|B| \leq 16c|A'|^2$. To complete the proof, we notice that

$$|A'|^2 = \sum_{x \in A' - A'} r^-_{A'}(x) = |B| + \sum_{x \in A' - A'} r^-_{A'}(x) \cdot A_{c\delta}(x).$$

□

Proof of Theorem 1.1. We apply Lemma 1.2 with the set A and $c = 1/128$, to obtain a set $A^* \subset A$ such that $|A^*| \geq \delta|A|/3$ and A^* satisfies (2). We also set $P = \left\{(x, y) \in (A^*)^2 : r^-_A(x - y) \geq \delta|A|/128\right\}$. By (2), P consists of at least $7|A^*|^2/8$ pairs.

Let A' be the set of elements of A^* that are the first coordinate of at least $3|A^*|/4$ pairs of P. It can be easily verified that $|A'| \geq |A^*|/2$. For any $a, b \in A'$, there are at least $|A^*|/2$ elements $y \in A^*$ such that both (a, y) and (b, y) are in P. For every such y we have $a - b = (a - y) - (b - y)$. That is, $a - b$ has at least $\frac{|A^*|}{2} \cdot \left(\frac{\delta|A|}{128}\right)^2 = \frac{\delta^2|A^*||A|^2}{2^{15}}$ representations of the form $(a_1 - a_2) - (a_3 - a_4)$ with $a_1, a_2, a_3, a_4 \in A$.

By summing the number of above representations for every difference in $|A' - A'|$, we get

$$\frac{\delta^2|A^*||A|^2}{2^{15}}|A' - A'| \leq |A|^4,$$

or

$$|A' - A'| \leq \frac{2^{15}|A|^2}{|A^*|\delta^2} = O\left(\frac{(|A'|/\delta)^2}{|A'|\delta^2}\right) = O\left(\frac{|A'|}{\delta^4}\right).$$

□

2 Alternative formulations of the theorem

After proving one formulation of the Balog-Szemerédi-Gowers theorem, we move to study alternative formulations. We begin by introducing the concept of *partial sum sets*. Consider a set $A \subset \mathcal{G}$ and a set of pairs $P \subset A \times A$. The partial sum set and partial difference set that are defined by P are

$$A \overset{P}{+} A = \{a + a' : \, a, a' \in A \text{ and } (a, a') \in P\},$$
$$A \overset{P}{-} A = \{a - a' : \, a, a' \in A \text{ and } (a, a') \in P\}.$$

Let us revisit the aforementioned example of a set with a large additive energy and a large sum set: $A = B \cup R$ where B is an arithmetic progression of size $|A|/2$ and R is a random set of size $|A|/2$. Recall that $|A + A| = \Theta(|A|^2)$. However, there exists a set of pairs $P \subset A^2$ of size $\Theta(|A|^2)$ such that $|A \overset{P}{+} A| = \Theta(|A|)$. Specifically, this is the set $P = B^2$.

Theorem 1.1 implies that if a set A has a large additive energy then it contains a large subset with a lot of structure (i.e., with a small sum set). A similar conclusion holds if there exists a large $P \subset A^2$ with $|A \overset{P}{+} A| = \Theta(|A|)$.

Lemma 2.1. *Given $A \subset \mathcal{G}$, the following statements are equivalent (specifically, each statement implies every other statement for some constants c_i).*
(i) $E^+(A) \geq c_1|A|^3$.
(ii) There exists $A' \subset A$ such that $|A'| \geq c_2|A|$ and $|A' - A'| \leq c_3|A'|$.
(iii) There exists $P \subset A^2$ such that $|P| \geq c_4|A|^2$ and $|A \overset{P}{-} A| \leq c_5|A|$.

Proof. By Theorem 1.1, statement (i) implies statement (ii) (with $c_2 = c_1$ and $c_3 = c_1^{-5}$). Statement (ii) implies statement (iii) by setting $P = A' \times A'$ (so $c_4 = c_2^2$ and $c_5 = c_3$). To show that statement (iii) implies statement (i), we use a variant of the standard Cauchy-Schwarz argument:

$$E^+(A) \geq \sum_{x \in A \overset{P}{-} A} |\{(a, a') \in P : \ a - a' = x\}|^2$$

$$\geq \frac{\left(\sum_{x \in A \overset{P}{-} A} |\{(a, a') \in P : \ a - a' = x\}|\right)^2}{|A \overset{P}{-} A|} \geq \frac{c_4^2 |A|^4}{c_5 |A|} = \frac{c_4^2 |A|^3}{c_5}.$$

□

So far we focused on difference sets and on results that involve a single set A. However, similar claims apply for sum sets and for problems that involve two sets. As an example we present the following theorem of Sudakov, Szemerédi, and Vu [2].

Theorem 2.2. *Let $A, B \subset \mathcal{G}$ with $|A| = |B| = n$, and let $P \subset A \times B$ satisfy $|P| = \delta n^2$ and $|A \overset{P}{+} B| = cn$. Then there exist subsets $A' \subset A$ and $B' \subset B$, each of size at least $\delta^2 n/16$, such that*

$$|A' + B'| \leq \frac{2^{12} c^3 n}{\delta^5}.$$

To prove the theorem, we rely on the following graph-theoretic lemma.

Lemma 2.3. *Let $G = (V \cup U, E)$ be a bipartite graph with $|V| = |U| = n$ and $|E| = \delta n^2$ for some $\delta > 0$. Then there exist subsets $V' \subset V$ and $U' \subset U$, each of size at least $\delta^2 n/16$, such that for every $v \in V'$ and $u \in U'$ there exist at least $2^{-12}\delta^5 n^2$ paths of length three between v and u in G.*

Proof. We first remove from U vertices that have a small degree. The average degree of a vertex in U is $|E|/|U| = \delta n$. We remove from U every vertex of degree at most $\delta n/2$. Since there are fewer than n such vertices, we remove fewer than $\delta n^2/2$ edges, so more than $\delta n^2/2$ edges remain in E.

We uniformly choose a vertex $v \in V$. Since more than $\delta n^2/2$ edges remain in E, we get that[1]

$$\mathbb{E}[|N(v)|] = \sum_{v \in V} \frac{|N(v)|}{|V|} = \frac{|E|}{|V|} > \frac{\delta n}{2}.$$

[1]Recall that $N(v)$ is the set of neighbors of v in the graph.

We say that the pair $(u, u') \in U^2$ is *bad* if $|N(u) \cap N(u')| < \delta^3 n/128$. For every bad pair (u, u') and a uniformly chosen $v \in V$, we have

$$\Pr\left[u \in N(v) \text{ and } u' \in N(v)\right] < \frac{\delta^3 n/128}{n} = \delta^3/128.$$

Let B_v be the set of bad pairs in $N(v)^2$. Since $|U| \le n$, there are at most $\binom{n}{2}$ bad pairs in U. By linearity of expectation

$$\mathbb{E}[|B_v|] < \binom{n}{2} \cdot \delta^3/128 < \frac{\delta^3 n^2}{256}.$$

Denote by $U_{B,v}$ the set of vertices of U that participate in at least $\delta^2 n/32$ pairs of B_v. Notice that

$$\mathbb{E}[|U_{B,v}|] \le \frac{2\mathbb{E}[|B_v|]}{\delta^2 n/32} < \frac{\delta n}{4}.$$

Finally, let $U_v = N(v) \setminus U_{B,v}$. That is, U_v is the set of neighbors of v that form fewer than $\delta^2 n/32$ bad pairs with other neighbors of v. By linearity of expectation

$$\mathbb{E}[|U_v|] = \mathbb{E}[|N(v)|] - \mathbb{E}[|U_{B,v}|] > \frac{\delta n}{2} - \frac{\delta n}{4} = \frac{\delta n}{4}.$$

Since this is the expected value for a uniformly chosen $v \in V$, there must exist $v \in V$ for which $|U_v| > \delta n/4$. We consider such a vertex v, set $U' = U_v$, and set

$$V' = \{v' \in V : |N(v') \cap U'| \ge \delta^2 n/16\}.$$

Consider two vertices $v' \in V'$ and $u \in U'$. By definition v' has at least $\delta^2 n/16$ neighbors in U', and fewer than $\delta^2 n/32$ of those form a bad pair with u. That is, U' contains more than $\delta^2 n/32$ vertices that are neighbors of v' and do not form a bad pair with u. For every such vertex u', there are at least $\delta^3 n/128$ vertices $v'' \in V$ such that $u, u' \in N(v'')$. Thus, the number of paths of the form (v', u', v'', u) is at least

$$\frac{\delta^2 n}{32} \cdot \frac{\delta^3 n}{128} = \frac{\delta^5 n^2}{2^{12}}.$$

It remains to prove that V' is not too small. Since every remaining vertex of U is of degree at least $\delta n/2$, the number of edges between V and U' satisfies

$$|E(V, U')| \ge \delta n/2 |U'| \ge \delta^2 n^2/8.$$

Since every vertex of $V \setminus V'$ corresponds to fewer than $\delta^2 n/16$ such edges, we have $|E(V', U')| > \delta^2 n^2/16$. Since the maximum degree of a vertex is n, we obtain that $|V'| > \delta^2 n/16$. □

Proof of Theorem 2.2. We build a bipartite graph $G = (A \cup B, E)$ (i.e., one side contains a vertex for every element of A and the other a vertex for every element of B) with $E = P$. This graph satisfies the assumptions of Lemma 2.3. By this lemma, there exist subsets $A' \subset A$ and $B' \subset B$, each of size at least $\delta^2 n/16$, such that for every $a \in A'$ and $b \in B'$ there exist at least $2^{-12}\delta^5 n^2$ paths of length three between a and b.

Consider $a \in A'$ and $b \in B'$ and a path of length three (a, b', a', b), where $a' \in A$ and $b' \in B$. We have

$$a + b = (a + b') - (b' + a') + (a' + b).$$

Since $(a, b'), (b', a'), (a', b) \in E$, we get that $a+b', b'+a', a'+b \in A \overset{P}{+} B$. That is, we can represent $a + b$ as a sum of three elements from $A \overset{P}{+} B$, so that this sum corresponds to the path (a, b', a', b). Since there are at least $2^{-12}\delta^5 n^2$ such paths, this is also a lower bound for the number of solutions of $a+b = x+x'+x''$ where $x, x', x'' \in A \overset{P}{+} B$. Since $|A \overset{P}{+} B| = cn$, there are $c^3 n^3$ triples $(x, x', x'') \in (A \overset{P}{+} B)^3$. This implies

$$|A' + B'| \le \frac{c^3 n^3}{2^{-12}\delta^5 n^2} = \frac{c^3 2^{12} n}{\delta^5},$$

as asserted. □

Other interesting variants of the Balog-Szemerédi-Gowers exist. For example, there are asymmetric variants for the case of two sets A and B of distinct sizes (e.g., see [3, Section 2.6]).

References

[1] T. Schoen, New bounds in Balog-Szemerdi-Gowers theorem, *Combinatorica*, to appear.

[2] B. Sudakov, E. Szemerédi, and V. H. Vu, On a question of Erd˝s and Moser, *Duke Mathematical Journal* **129** (2005), 129–155.

[3] T. Tao and V. H. Vu, *Additive combinatorics*, Cambridge University Press, 2006.