

5. AN APPROXIMATE SUBRING A HAS SMALL $|A \cdot A + A \cdot A|$

Today we will fill the gap left from last time, namely that if a subset $A \subseteq F$ has small sumset $A + A$ and product set $A \cdot A$ has small sum-product set $A \cdot A + A \cdot A$. This may seem clear given the notation, but it actually takes a bit of work to prove. The key ingredient is Tim Gowers' quantitative version of a theorem of Balog-Szemerédi (see [1]). According to Gowers in [2], the original proof of Balog-Szemerédi may be traced through to find bounds on the constants at play, but the bounds are poor and of little use in practice. Gowers found a nice, elementary proof (requiring only Hölder's inequality, a proof of which we include in the appendix for the sake of completeness) that not only reproves the Balog-Szemerédi theorem, but also gives good information on the constants involved.

Recall that during weeks 1 and 2 we found a polynomial P such that, in the situation of Theorem 0.1, $P(A, \dots, A) = F$. Last week, we showed in full detail that if the sum-product set $A \cdot A - A \cdot A$ is small, then $P(A, \dots, A)$ is small in such a way that lead to a contradiction. Thus we were able to conclude that the BKT theorem was true. The missing piece is this (recall that the \lesssim notation means $O(\cdot)$ where the implied constant is independent of the argument):

Theorem 5.1. *Let $A \subseteq F$ such that $|A + A|, |A \cdot A| \leq K|A|$ for some constant K . Then there is a subset $A' \subseteq A$ with $|A'| \approx |A|$ and $|A' \cdot A' + A' \cdot A'| \lesssim |A'|$.*

The careful reader will now notice that, unless there is a stronger result than this, we must have been lying a little bit last week in the proof of BKT. So far, I have always been promising that if A is an approximate subring, then its product-sum set is small, but this is not necessarily true. In general, we must pass to a large subset, but this does not affect the proof of BKT since the polynomial P depends only on δ ; in particular, it is independent of the subset A . Thus, although it's not the result one might have expected, it is sufficient for our purposes.

Before we prove Theorem 5.1, we will state and prove Gowers' result. We let G denote some (additive) abelian group. The notation that we use is Gowers', and as he states in [2], it is non-standard, so we spend a minute on it here. Let $f, g : G \rightarrow \mathbb{Z}$, and define a convolution operator $*$ by

$$(f * g)(x) = \sum_{x=s-t} f(s)g(t)$$

We will identify a set A with its characteristic function χ_A , so then

$$A * A(x) = \sum_{x=s-t} \chi_A(s)\chi_A(t)$$

is the number of ways that an element $x \in G$ may be written as a difference in $A - A$. Similarly, for each x , the quantity $(A * A(x))^2$ is the number of quadruples $(a, b, c, d) \in A^4$ such that $a - b = x = c - d$, so $\|A * A(x)\|_2^2$ is the number of quadruples $(a, b, c, d) \in A^4$ such that $a - b = c - d$. Now we state Gowers' result ([2], Proposition 12):

Proposition 5.2. *Let $A \subseteq G$, where G is an abelian group, $|A| = m$ is finite, and suppose that $\|A * A\|_2^2 \geq c_0 m^3$ for some constant c_0 . Then there is a constant C (depending only on c_0) and a subset $A' \subseteq A$ such that $|A'| \geq \frac{m}{C}$ and $|A' - A'| \leq Cm$. In fact, for all $a, a' \in A'$, there are at least m^7/C solutions (a_1, \dots, a_8) to the equation*

$$a - a' = (a_1 - a_2) - (a_3 - a_4) - ((a_5 - a_6) - (a_7 - a_8))$$

Proof. The function $f(x) = A * A(x) : G \rightarrow \mathbb{Z}$ is nonnegative and it is easy to see that f satisfies:

- $\|f\|_\infty \leq m$
- $\|f\|_1 = m^2$

Therefore $f(x) \geq c_0m/2$ for at least $c_0m/2$ values of x ; otherwise, there would exist a subset $S \subseteq G$, $|S| < c_0m/2$, such that $f(x) \geq c_0m/2$ if and only if $x \in S$, so

$$\begin{aligned} \|f\|_2^2 &= \sum_{x \in S} f(x)^2 + \sum_{x \notin S} f(x)^2 \\ &\leq |S| \|f\|_\infty^2 + \frac{c_0m}{2} \sum_{x \in G} f(x) \\ &< \frac{c_0m}{2} m^2 + \frac{c_0m}{2} m^2 = c_0m^3 \end{aligned}$$

which contradicts the assumption on $\|f\|_2^2$. Call $x \in G$ a **popular difference** if $f(x) \geq c_0m/2$ and define a graph Γ to have vertex set equal to A and an edge between $a, b \in A$ if and only if $a - b$ is a popular difference.

Now we claim that the average degree in Γ is at least $\frac{c_0^2m}{4}$. To see this, first we note that by above, there are at least $c_0m/2$ values of x such that x is a popular difference. But $f(x)$ is the number of ways that x may be written as $x = a - a'$ for $a, a' \in A$, so there are at least $c_0m/2$ popular differences in $A - A$. Therefore, there are at least $c_0m/2$ distinct differences $a - a' \in A - A$ such that $f(a - a') \geq c_0m/2$, so for each of these pairs there are at least $c_0m/2$ pairs (x, y) such that $x - y = a - a'$, each of which corresponds to a distinct edge in Γ , so there are at least $(c_0m/2)^2$ distinct edges in Γ . Thus the claim follows.

By the claim, there are at least $c_0^2m/8$ vertices with degree at least $c_0^2m/8$. Set $\delta = c_0^2/8$ and let a_1, \dots, a_n be those vertices with high degree (where $n \geq \delta m$), and let $N_1(a_i)$ denote the 1-neighborhood in Γ of a_i . By a technical combinatorial lemma, which we relegate to the appendix, there is a subset $A' \subseteq \{a_1, \dots, a_n\}$ such that $|A'| \geq \delta^5 n / \sqrt{2}$ and $|N_1(a_i) \cap N_1(a_j)| \geq \delta^2 m / 2$ for at least 90% of the pairs $(a_i, a_j) \in (A')^2$. Set $\alpha = \delta^2 / \sqrt{2}$, so $|A'| \geq \alpha m$.

Define a new graph Γ' with vertex set A' and edges (a_i, a_j) whenever $|N_1(a_i) \cap N_1(a_j)| \geq \delta^2 m / 2$. By the above, the average degree is $9/10|A'|$, so at least $4/5|A'|$ vertices have degree at least $4/5|A'|$. Let A'' be all such vertices. First observe that

$$|A''| \geq \alpha m = \frac{\delta^6}{\sqrt{2}} m = \left(\frac{c_0^2}{8}\right)^6 \frac{\sqrt{2}}{2} m$$

so A'' has the desired cardinality. Finally, we claim that A'' has small difference set $A'' - A''$. To see this, let $a_i, a_j \in A''$. By the definition of Γ' , the degrees of a_i and a_j are at least $4/5|A'|$ in Γ' , so there exist at least $3/5|A'|$ points $a_k \in A'$ connected to both a_i and a_j in Γ' . By the definition of Γ' , $|N_1(a_i) \cap N_1(a_k)|$ and $|N_1(a_j) \cap N_1(a_k)|$ are at least $\delta^2 m / 2$. Now suppose that $b \in N_1(a_i) \cap N_1(a_k)$. Then $(a_i, b), (a_k, b) \in E(\Gamma)$, the edge set of Γ , so $a_i - b$ and $a_k - b$ are popular differences. Thus, there are at least $(c_0m/2)^2$ ways of writing

$$a_i - a_k = (a_i - b) - (a_k - b) = (p - q) - (r - s)$$

for $(p, q, r, s) \in A^4$, and a similar statement is true for $a_j - a_k$. By unravelling definitions, putting together inequalities, and summing over the at least $3/5|A'|$

points a_k that are connected to both a_i and a_j , one finds that there are at least

$$(3/5)|A'|\delta^4 c_0^4 m^6 / 64 \geq \alpha \delta^4 c_0^4 m^7 / 120$$

ways of writing $a_i - a_j$ as an element of $4A - 4A$. \square

Exercise 5.1. *Provide the details for the end of the proof of Proposition 5.2.*

We will use a slight generalization of the proposition to prove the main result for today. See [3] for an outline of a proof:

Lemma 5.3. *Let $A, B \subseteq G$, $|A| = |B|$, and $|A+B| \leq K|A|$. Then there are subsets $A' \subseteq A$ and $B' \subseteq B$ such that $|A'| \geq C|A|$ and $|B'| \geq C|B|$ for some constant C (depending only on K) and, for fixed $a' \in A'$ and $b' \in B'$, there are at least $|A|^5/C$ solutions to*

$$a' - b' = (a_1 - b_1) - (a_2 - b_2) + a_3 - b_3$$

with $a_i \in A$ and $b_i \in B$.

Exercise 5.2. *Either modify the proof of Proposition 5.2 or expand Tao's argument from [3] to prove the lemma.*

Proof (of Theorem 5.1). The lemma implies that there are subsets $C, D \subseteq A$ such that $|C|, |D| \approx |A|$ and every element of $C - D$ has approximately $|A|^5$ representations of the form

$$c - d = (a_1 - a_2) - (a_3 - a_4) + (a_5 - a_6)$$

with $a_i \in A$. In anticipation of using the full hypothesis that A is an approximate subring, we multiply by an arbitrary element of $A \cdot A \cdot A/A \cdot A$ to find approximately $|A|^5$ representations of the form

$$(13) \quad (b_1 - b_2) - (b_3 - b_4) + (b_5 - b_6)$$

(with $b_i \in A \cdot A \cdot A \cdot A/A \cdot A$) for elements in $(C - D) \cdot A \cdot A \cdot A/A \cdot A$. By removing 0 from A (if necessary), we may apply the multiplicative form of sumset estimates (Theorem 3.3), we know that $|A \cdot A \cdot A \cdot A/A \cdot A| \leq K^6|A|$. Therefore, since the total number of possible representations of the form (13) is $|A|^6$, we must have

$$(14) \quad |(C - D) \cdot A \cdot A \cdot A/A \cdot A| \approx |A|$$

Now we refine C and D in order to find the desired subset $A' \subseteq A$. Since $C \cdot D \subseteq A \cdot A$ and A is an approximate subring, we know that $|C \cdot D| \approx |A|$. Therefore $|C \cdot D| \approx |C| \approx |D|$, so by the multiplicative form of Lemma 5.3, there are subsets $C' \subseteq C$ and $D' \subseteq D$ with $|C'| \approx |C| \approx |A| \approx |D| \approx |D'|$ such that every element of $C' \cdot D'$ has approximately $|A|^5$ representations of the form

$$\frac{c_1 d_1 c_3 d_3}{c_2 d_2}, c_i \in C, d_i \in D$$

with $c_i \in C$ and $d_i \in D$.

Let $c, c' \in C'$ and $d, d' \in D'$ be arbitrary. Then by the pigeonhole principle, there exist $c_2 \in C$ and $d_2 \in D$ such that there are approximately $|A|^3$ solutions to

$$cd = \frac{c_1 d_1 c_3 d_3}{c_2 d_2}$$

We may rewrite this as

$$cd - c'd' = x_1 - x_2 + x_3 - x_4$$

where

$$\begin{aligned} x_1 &= \frac{(c_1 - d')d_1c_3d_3}{c_2d_2} \\ x_2 &= \frac{d'(c' - d_1)c_3d_3}{c_2d_2} \\ x_3 &= \frac{d'c'(c_3 - d_2)d_3}{c_2d_2} \\ x_4 &= \frac{d'c'd_2(c_2 - d_3)}{c_2d_2} \end{aligned}$$

For fixed c', d', c_2, d_2 , it is not hard to see that the map sending (c_1, c_3, d_1, d_3) to (x_1, x_2, x_3, x_4) is injective, and hence a bijection onto its image. Therefore, since all the x_j lie in $(C - D) \cdot A \cdot A \cdot A / A \cdot A$, we thus have approximately $|A|^3$ ways of representing $cd - c'd' \in C' \cdot D' - C' \cdot D'$ in the form $x_1 - x_2 + x_3 - x_4$. By 14 (and an argument similar to that preceding 14), we have

$$(15) \quad |C'D' - C'D'| \lesssim |A|$$

Therefore we surely have $|C'D'| \leq |C'D' - C'D'| \lesssim |A| \approx |C'|$, and thus by the multiplicative form of Lemma 5.3⁶ there are large subsets $C'' \subseteq C'$ and $D'' \subseteq D'$ such that

$$(16) \quad |C''/D''| \lesssim |C'|$$

Finally, we consider the map $\pi : C'' \times D'' \rightarrow C''/D''$ defined in the obvious way, $(x, y) \mapsto x/y$. Since we have the bound (16), the pigeonhole principle guarantees the existence of an $x/y \in C''/D''$ such that

$$|\pi^{-1}(x/y)| \geq \frac{|C''||D''|}{|C''/D''|} \approx |A|$$

But for all $(c, d) \in \pi^{-1}(x/y)$, we know $c = d(x/y)$. Thus

$$|C'' \cap (D'' \cdot (x/y))| = |\pi^{-1}(x/y)| \approx |A|$$

Now we set $A' := C'' \cap (D'' \cdot (x/y))$, so $|A'| \approx |A|$, and, by translation invariance,

$$|A' \cdot A' - A' \cdot A'| \leq |C' \cdot D' - C' \cdot D'| \lesssim |A|$$

which completes the proof. \square

So we have seen the original proof of the BKT theorem from first principles in complete detail⁷ and I hope that it has provided a good introduction to arithmetic combinatorics. The main tools that we used were the Cauchy-Davenport inequality, sum-set (and product-set) estimates, and a quantitative version of the Balog-Szemerédi Theorem. We also used a few clever lemmas of Imre Ruzsa as well as many impressive arguments of Bourgain, Katz, and Tao. All of the proofs presented here are either from Tao's notes [3], the BKT paper [1], or Gowers' paper on arithmetic progressions of length four [2]. One of the curiosities of the complete proof of BKT, as presented in this course, is that we never had to leave the world of cardinalities of sets; for example, we never applied deep structure theorems like

⁶Tao et al. say this follows from product-set estimates, i.e. the multiplicative form of sumset estimates (our Theorem 3.3), but I can't see it.

⁷See the appendix for more on "complete detail".

Freiman's theorem, although they may have been helpful at times. During the next few weeks, we will look briefly at some generalizations and applications of BKT.

References.

- [1] A. Balog and E. Szemerédi. A statistical theorem of set addition. *Combinatorica*, 14:263–268, 1994.
- [2] W.T. Gowers. A new proof of szemerédi's theorem of arithmetic progressions of length four. *Geom. Funct. Anal.*, 8:529–551, 1998.
- [3] Terry Tao. Math 254a: Some highlights of arithmetic combinatorics, 2003.