

Szemerédi's Theorem on Three Term Progressions, at a Glance

April 3, 2007

1 Introduction

Here we will give only the very briefest sketch of the ideas in Szemerédi's proof that for every $0 < \theta \leq 1$ and N sufficiently large, any subset

$$S \subseteq [N] := \{1, \dots, N\}, \quad |S| \geq \theta N,$$

contains a three-term arithmetic progression. For those wanting a more complete exposition, see Akos Magyar's website for a very nice writeup.

2 Sketch of the Proof

In virtually all combinatorial proofs of results of this sort, the idea is show that if S has no three-term progressions, then there is a long arithmetic progression on which “ S has increased density”, by this I mean, we try to find an arithmetic progression P such that

$$|S \cap P| \geq (\theta + \epsilon)|P|,$$

where, say, $N' = |P| > \log N$ or something. Obviously, this reduces the problem to showing that sets $S' \subset [N']$ of density at least $\theta + \epsilon$ have three-term progressions. Then, we just iterate. We now state a key principle Szemerédi applies to find such P .

2.1 A Basic Density Increment Principle

Density Increment Principle. Let T be the complement of S inside $[N]$. Suppose T contains a subset U with δN elements, which is approximately translation invariant, in the sense that there exists t such that

$$|U + \{0, t\}| \sim |U|.$$

Then, there exists a long arithmetic progression P where

$$|S \cap P| \gtrsim (1 - \delta)^{-1} \theta |P|. \tag{1}$$

The proof of this is simple: First, $[N]$ can be covered by $\sim N$ arithmetic progressions of N' terms (recall $N' = |P|$) having common difference t , and all but a few points of $[N]$ each lie in N' of these progressions, at least for t and N' small enough. Now, by the translation-invariance of U one deduces that many of these progressions, about δN of them, lie entirely in $U \subseteq T$, and therefore contain no points of S . So, the remaining $\sim (1 - \delta)N$ progressions completely cover S , and each point of S (except a few) lies in exactly N' of them. We are done, since this quickly implies that (1) holds for one of these progressions P .

2.2 The Construction of U and the “Szemerédi Cube Lemma”

We consider the elements of S in each of three intervals

$$T_1 = (0, N/6] \cap S, \quad T_2 = (5N/12, N/2] \cap S, \quad T_3 = (2N/3, N] \cap S.$$

We assume

$$|T_1| \sim N/6, \quad |T_2| \sim \theta N/12, \quad \text{and} \quad |T_3| \sim \theta N/3,$$

else S has increased density on one of these three intervals, in which case we iterate.

Now, we will apply the “Density Increment Principle” to the set T_3 , by constructing the requisite set U inside of T_3 . The way we find U is to find an affine cube sitting in T_2 , and we do this using the following:

Szemerédi Cube Lemma. Each integer subset $A \subseteq [x, x + M]$ with $|A| \geq \theta M$ contains a sumset

$$a + \{0, x_1\} + \{0, x_2\} + \cdots + \{0, x_d\}, \text{ where } d > c_\theta \log \log M.$$

We will not prove this lemma here, suffice it to say that an easy induction argument does the job.

Now let C be a Szemerédi cube in T_2 of dimension $d \gg \log \log N$. If S has no three-term progressions, then

$$U' = 2C - T_1 \text{ is disjoint with } T_3, \text{ yet lies in } [2N/3, N].$$

We find our U by consider the following chain of sets:

$$\begin{aligned} U_1 &= 2a - T_1 \subseteq U' \\ U_2 &= U_1 + \{0, 2x_1\} \subseteq U' \\ U_3 &= U_2 + \{0, 2x_2\} \subseteq U' \\ &\vdots \\ U_{d+1} &= U_d + \{0, 2x_d\} \subseteq U'. \end{aligned}$$

We have

$$\theta N/6 \sim |U_1| \leq |U_2| \leq \cdots \leq |U_{d+1}| \leq |U'| \leq N/3.$$

Clearly we will have that for some index i ,

$$|U_i| \sim |U_{i+1}| = |U_i + \{0, 2x_i\}|,$$

and we are done on letting $U = U_i$ and $t = 2x_i$.