Szemerédi's proof of Roth's theorem that $r_3(n) = o(n)$

Martin Klazar[1]

April 18, 2013

I present Szemerédi's combinatorial proof [4] of Roth's theorem [2, 3] on arithmetic progressions of length three. My motivation to write it up was the beauty of the whole argument, as well as my recent realization that my understanding of it contains a (small) gap. I comment on this gap and a minor innovation in the proof at the end.

$\mathbb{N} = \{1, 2, \dots\}$ and $[n] = \{1, 2, \dots, n\}$ for $n \in \mathbb{N}$. AP is an abbreviation for 'arithmetic progression'. This is a subset of $\mathbb{N}$ of the form $\{a, a+d, a+2d, \dots, a+(m-1)d\}$ where $a, m, d \in \mathbb{N}$; in particular, always $d > 0$. $|X|$ denotes cardinality of the set X. For $X \subset \mathbb{N}$ and $a \in \mathbb{N}$, we use notation $X + a = \{x + a \mid x \in X\}$.

Theorem (Roth, 1952). *If $r_3(n)$ is the maximum size of a subset of $[n]$ containing no AP $\{a, a+d, a+2d\}$, then*

$$r_3(n) = o(n),\ n \to \infty\ .$$

Equivalently: for every $\delta > 0$ there is an $n_0 \in \mathbb{N}$ such that if $n > n_0$ and $X \subset [n]$ with $|X| > \delta n$, then X contains a 3-term AP.

Let $\delta \in (0, 1]$ be a real number. A *δ-sequence* is an infinite sequence of pairs (X_i, n_i), $i = 1, 2, \dots$, where $0 < n_1 < n_2 < \dots$ are integers, $X_i \subset [n_i]$ are subsets and, for $i \to \infty$,

$$\frac{|X_i|}{n_i} \to \delta\ (> 0)\ .$$

We restate Roth's theorem in terms of δ-sequences.

Proposition. *Every δ-sequence (X_i, n_i) contains a 3-term AP:*

$$X_i \supset \{a, a+d, a+2d\}$$

for some i (equivalently, for every $i > i_0$).

We prove Roth's theorem in the form of the Proposition. The proof uses three lemmas.

A set $X \subset \mathbb{N}$ *contains an l-cube*, $l \in \mathbb{N}$, if there exist positive integers $a_1, a_2, \dots, a_l$ and sets

$$\emptyset \ne Q_1 \subset Q_2 \subset \dots \subset Q_{l+1} = X \ \text{ with }\ Q_j + a_j \subset Q_{j+1}\ \text{ for } 1 \le j \le l\ .$$

[1] klazar@kam.mff.cuni.cz

Lemma 1. *Every δ-sequence (X_i, n_i) contains (i.e., X_i contains, for $i > i_0$) an l-cube for every $l \in \mathbb{N}$.*

Proof. Induction on l. For $l = 1$, X_i contains a 1-cube whenever $|X_i| \geq 2$ (if $X_i = \{a < b < \dots\}$, write $b = a + (b - a) = a + a_1$), which holds for any large i. Suppose that the lemma holds for $l \geq 1$ and every δ-sequence. We claim that for every δ-sequence (X_i, n_i) there exist subsets $Y_i \subset X_i$ and integers $b_i > 0$ such that (Y_i, n_i) is a $\delta^2/2$-sequence and $Y_i + b_i \subset X_i$ for every i. Then we apply induction on the sequence (Y_i, n_i), extend the l-cube in Y_i by $Q_{l+2} = X_i$ and $a_{l+1} = b_i$, and get an $(l + 1)$-cube in X_i.

To establish the claim, we set Y_i' to be the as of the pairs $a < b$ in X_i realizing the most popular distance $b - a$ between two elements of X_i, and set $b_i = b - a$ to be that distance. By the pigeon-hole, $|Y_i'| > \binom{|X_i|}{2}/n_i = \frac{1}{2}(|X_i|^2/n_i - |X_i|/n_i)$. Thus $|Y_i'|/n_i > \frac{1}{2}(|X_i|/n_i)^2 + O(1/n_i)$. Throwing away elements from Y_i' if necessary, we get Y_i with $|Y_i|/n_i \to \delta^2/2$ for $i \to \infty$. It is clear that for every i, $Y_i \subset X_i$ and $Y_i + b_i \subset X_i$. □

Note that for an l-cube in X_i some a_j may be as large as, say, $a_j > n_i/2$, but together we have $a_1 + a_2 + \cdots + a_l < n_i$.

A δ-sequence (X_i, n_i) is *saturated* if for every $\varepsilon > 0$ there is an m such that, for every i, if $A \subset [n_i]$ is an AP with $|A| \geq m$ then

$$\frac{|X_i \cap A|}{|A|} < \delta + \varepsilon \ .$$

If $X \subset [n]$ and $A = \{a, a + d, \dots, a + (m - 1)d\} \subset [n]$ is an AP, we set

$$X \,|\, A = \{j \in [m] \mid a + (j - 1)d \in X\} = (x \mapsto a + (x - 1)d)^{-1}(X \cap A) \ .$$

The *restriction* $X \,|\, A$ records the positions of the elements of X in the AP A. Note that $|X \,|\, A| = |X \cap A|$, $X \cap A$ is an AP if and only if $X \,|\, A$ is an AP and that one has this transitivity: if $B \subset [m]$ is another AP then $(X \,|\, A) \,|\, B = X \,|\, C$ where $C \subset A$ is the unique AP with $C \,|\, A = B$.

Lemma 2. *For every δ-sequence (X_i, n_i) there exist indices $i_1 < i_2 < \dots$ and APs $A_j \subset [n_{i_j}]$ with lengths m_j such that $m_1 < m_2 < \dots$ and*

$$(Y_j, m_j) = (X_{i_j} \,|\, A_j, m_j)$$

is a saturated δ'-sequence with $\delta' \geq \delta$.

Proof. If (X_i, n_i) is saturated we do nothing and set $i_j = j$, $A_j = X_j$ and $\delta' = \delta$. Else there exist a $\delta_0 > \delta$, indices $i_1 < i_2 < \dots$ and APs $A_j \subset [n_{i_j}]$ such that $|A_1| < |A_2| < \dots$ and $|X_{i_j} \cap A_j|/|A_j| > \delta_0$ for every j. Let δ' be the supremum of all δ_0 with this property. By the definition of δ' there exist indices $i_1 < i_2 < \dots$ and APs $A_j \subset [n_{i_j}]$ such that $|A_1| < |A_2| < \dots$ and

$$\frac{|X_{i_j} \cap A_j|}{|A_j|} > \delta' - \frac{1}{j}$$

for every $j \in \mathbb{N}$. This is the sequence of indices and APs we seek. By the maximality of δ', $|X_{i_j} \cap A_j|/|A_j| = |X_{i_j} \mid A_j|/|A_j| \to \delta'$ as $j \to \infty$. Also, $(X_{i_j} \mid A_j, |A_j|)$ is saturated, for else the above mentioned transitivity would give for the original δ-sequence indices and APs producing a value δ_0 with $\delta_0 > \delta'$, contradicting the definition of δ'. □

By Szemerédi's theorem, any δ-sequence contains an l-term AP for any l; thus Lemma 2 holds in fact with $|Y_j|/m_j = 1$ for every j.

If $X \subset \mathbb{N}$ and $d \in \mathbb{N}$, the *d-decomposition of* X is the unique expression of X as a disjoint union

$$X = \bigcup_{j=1}^{r} A_j$$

of nonempty APs A_j with the same common difference d and the property that, for every j, both $\min A_j - d \notin X$ and $\max A_j + d \notin X$. We obtain it by intersecting X with the d congruence classes modulo d, and then partitioning each nonempty intersection into maximal intervals of consecutive elements.

Lemma 3. *The d-decomposition $X = \bigcup_{j=1}^{r} A_j$ of X has the following properties.*

1. *The number of progressions is bounded by $r \leq |(X + d)\backslash X|$.*

2. *Let $m, n \in \mathbb{N}$ with $n \geq md$ and let $X \subset [n]$. Define $X' = \bigcup_{j=1}^{r} A'_j \subset X$ where each A'_j arises from A_j by deleting the first m and the last m elements ($A'_j \neq \emptyset$ iff $|A_j| > 2m$). Then in the d-decomposition of the complement*

$$[n]\backslash X' = \bigcup_{j=1}^{s} B_j$$

each AP B_j has length at least m.

Proof. 1. The mapping $A_j \mapsto \max A_j + d$ is an injection and goes from $\{A_1, \ldots, A_r\}$ to $(X + d)\backslash X$.

2. By the definition, B_j is a maximal interval of $C \cap ([n]\backslash X')$ where C is a mod d congruence class. If B_j is in $C \cap [n]$ followed or preceded by a nonempty A'_k, then B_j contains the first m or the last m elements of A_k and $|B_j| \geq m$. If there is no such A'_k then $B_j = C \cap [n]$ and we have $|B_j| \geq m$ due to the assumption that $n \geq md$. □

Szemerédi's proof of Roth's theorem. We prove that every δ-sequence (X_i, n_i) contains a 3-term AP. By Lemma 2 and the observations on $X \mid A$, we may assume that (X_i, n_i) is saturated. We split each $[n_i]$ into three intervals (which are also APs)

$$[n_i] = I_i \cup J_i \cup K_i = [\lfloor n_i/4 \rfloor] \cup [\lfloor n_i/4 \rfloor + 1, \lfloor n_i/2 \rfloor] \cup [\lfloor n_i/2 \rfloor + 1, n_i] .$$

I_i, J_i and K_i have respective lengths, up to errors $O(1)$, $n_i/4$, $n_i/4$ and $n_i/2$. We set

$$U_i = X_i \cap I_i,\ V_i = X_i \cap J_i \ \text{ and } \ W_i = X_i \cap K_i \ .$$

Since (X_i, n_i) is a δ-sequence and is saturated, for large i we have

$$|U_i|, |V_i| \geq \delta n_i/5 = \delta n_i/4 - \delta n_i/20 \ ,$$

because $|U_i|, |V_i| < \delta n_i/4 + \delta n_i/60$, $|W_i| < \delta n_i/2 + \delta n_i/60$ and $|U_i| + |V_i| + |W_i| = |X_i| > \delta n_i - \delta n_i/60$ for large i. We show that for large i there is a 3-term AP $u, v = u + d, w = u + 2d$ in X_i with $u \in U_i$, $v \in V_i$ and $w \in X_i$. As $u + w = 2v$, this is equivalent with finding such elements $u \in U_i$ and $v \in V_i$ that $2v - u$ is in X_i. Note that $2v - u \in [n_i]$ for every $v \in J_i$ and $u \in I_i$.

Using saturatedness of (X_i, n_i), we fix a large $m \in \mathbb{N}$ such that

$$\frac{|X_i \cap A|}{|A|} < \delta + \delta^2/40$$

whenever $A \subset [n_i]$ is an AP with $|A| \geq m$. Then we fix a large $l \in \mathbb{N}$ such that $2m/l < \delta/10$. By Lemma 1, for each large i the set V_i contains an $(l+2m)$-cube:

$$\emptyset \neq Q_1 \subset Q_2 \subset \ldots \subset Q_{l+2m+1} = V_i \ \text{ with } \ Q_j + a_j \subset Q_{j+1} \ \text{ for } 1 \leq j \leq l+2m \ ,$$

for some sets Q_j and positive integers $a_1, a_2, \ldots, a_{l+2m}$ (for simplicity of notation we do not mark explicitly their dependence on i). Now

$$a_1 + a_2 + \cdots + a_{l+2m} < n_i \ ,$$

and thus $a_j > n_i/2m$ only for at most $2m$ indices j. Without loss of generality, the big a_js are the last ones. Hence

$$2m/l < \delta/10 \ \text{ and } \ 2a_j m \leq n_i,\ j = 1, 2, \ldots, l \ .$$

We define

$$D_j = 2Q_j - U_i = \{2v - u \mid v \in Q_j, u \in U_i\},\ 1 \leq j \leq l + 1 \ .$$

Clearly,

$$D_1 \subset D_2 \subset \ldots \subset D_{l+1} \subset [n_i],\ D_j + 2a_j \subset D_{j+1} \ \text{ and } \ n_i \geq |D_j| \geq |D_1| \geq |U_i| \ .$$

It follows that $|D_{j+1} \backslash D_j| < n_i/l$ for some j, $1 \leq j \leq l$. We consider the $2a_j$-decomposition

$$D_j = \bigcup_{t=1}^{r} A_t$$

for this j, and the set

$$E_i = \bigcup_{t=1}^{r} A_t' \subset D_j \ ,$$

with A'_t obtained by deleting the first and last m elements of the AP A_t. Thus $|D_j \backslash E_i| \le 2mr$. By the two properties of d-decompositions in Lemma 3,

$$r \le |(D_j + 2a_j)\backslash D_j| \le |D_{j+1}\backslash D_j| < n_i/l \ \text{ and } [n_i]\backslash E_i = \bigcup_{t=1}^{s} B_t,\ |B_t| \ge m\ ,$$

where the B_t are disjoint APs with common difference $2a_j$. Each B_t has indeed length at least m because $2a_j m \le n_i$.

For large i, due to the selection of m and l, due to saturatedness of (X_i, n_i) and due to the fact that each $|B_t| \ge m$, we have

$$|E_i| \ge |D_j| - 2mr \ge |U_i| - (2m/l)n_i > \delta n_i/5 - \delta n_i/10 = \delta n_i/10$$

and

$$\begin{aligned}
|X_i \cap E_i| &= |X_i| - |X_i \cap ([n_i]\backslash E_i)| \\
&> \delta n_i - \delta^2 n_i/40 - \sum_{t=1}^{s} |X_i \cap B_t| \\
&> \delta n_i - \delta^2 n_i/40 - (\delta + \delta^2/40)\sum_{t=1}^{s} |B_t| \\
&= \delta n_i - \delta^2 n_i/40 - (\delta + \delta^2/40)(n_i - |E_i|) \\
&> \delta |E_i| - \delta^2 n_i/20 > \delta^2 n_i/10 - \delta^2 n_i/20 \\
&= \delta^2 n_i/20\ .
\end{aligned}$$

Hence $X_i \cap E_i \ne \emptyset$ and $w \in X_i \cap E_i$ for large i. Since $E_i \subset D_j = 2Q_j - U_i$ and $Q_j \subset V_i$, this means that $w = 2v - u \in X_i$ with $v \in V_i$ and $u \in U_i$, and we get the desired 3-term AP $\{u, v, w\} \subset X_i$. □

Remarks. I took the proof from the book [1] of Moreno and Wagstaff (it also contains Szemerédi's proof of Szemerédi's theorem). The small gap that I did not realize until recently is the missing of the reduction from $(l + 2m)$-cube to l-cube to purge large $2a_j$s. Without this purge, if we pick up a bad index j with $2a_j > n_i/m$, the final calculation does not work as all APs B_t are short. This gap is present in the rendering of the proof in the book [1] (and maybe elsewhere). Of course, once one realizes it, it is easy to fix it but it brings a humbling experience. The minor innovation of mine is the introduction of δ-sequences in the proof, which makes some arguments cleaner (seems to me) and reduces necessary calculations. The transition to saturated subsequence in Lemma 2 corresponds, in other renderings of Szemerédi's proof, to the step of proving by Fekete's lemma that $\lim_{n\to\infty} r_3(n)/n$ exists.

References

[1] C. J. Moreno and S. S. Wagstaff, *Sums of Squares of Integers*, Chapman & Hall/CRC, Boca Raton, FL, 2006.

[2] K. Roth, Sur quelques ensembles d'entiers, *C. R. Acad. Sci. Paris*, **234** (1952), 388–390.

[3] K. F. Roth, On certain sets of integers, *J. London Math. Soc.*, **28** (1953), 104–109.

[4] E. Szemerédi, On sets of integers containing no four elements in arithmetic progression, *Acta Math. Acad. Sci. Hungar.*, **20** (1969), 89–104.