ROTH'S THEOREM
THE COMBINATORIAL APPROACH

AKOS MAGYAR

We present Szemeredi's proof on the existence of 3-progressions in large sets, which appeared some 10 years later the Roth' proof, and contain elements of his general result on the existence of k-progressions.

The key observation again is to show that, if a set $A \subset [1, N]$, $|A| = \delta N$ does not contain any 3-progressions, then there is a "long' progression P on which the density of A increases to, say $\delta + \delta^2/20$. This can be iterated, and if N is very large, then after a large number k of iterations, the density of A would increase to $\delta_k > 1$ on a progression P_k, which is clearly impossible.

Some preparations are helpful before we start the actual proof.

Definition 1. *Let A, B be to finite sets of integers. The density of A in B is defined by*

$$\delta(A|B) = \frac{|A \cap B|}{|B|}$$

Proposition 1. *Let A, $B = B_1 \cup \ldots B_k$ be to finite sets such that B is partitioned into the sets B_i. Then*

$$\delta(A|B) = \frac{|B_1|}{|B|}\, \delta(A|B_1) + \ldots + \frac{|B_k|}{|B|}\, \delta(A|B_k)$$

and in particular: $\delta(A|B) \leq \max_i \delta(A|B_i)$

The proof of this is left as an exercise, however we prove a stronger version below.

Proposition 2. *Let $1 \leq M < N$. Let $A \subset [1, N]$ such that $|A| = \delta N$.*

Assume that there is a set $B \subset [1, N]$ such that: $A \subset B$ and $|B| \leq (1 - \delta/10)\, N$.

If $B = B_1 \cup \ldots \cup B_k$ is a partition, such that the number of parts $k \leq N/M$, then there is a part B_i such that:

$$\delta(A|B_i) \geq (\delta + \delta^2/20) \quad \text{and} \quad |B_i| \geq \delta^3/20\ M$$

Proof. Let us call a part B_i *small* if $|B_i| \leq \delta^3/20\ M$ and *large* otherwise. Note that

$$\delta(A|B) = \frac{|A|}{|B|} \geq \frac{\delta}{1 - \delta/10} \geq \delta + \delta^2/10$$

Also

$$\delta(A|B) = \sum_{\{i:B_i \text{ small}\}} \frac{|B_i|}{|B|}\,\delta(A|B_i) + \sum_{\{i:B_i \text{ large}\}} \frac{|B_i|}{|B|}\,\delta(A|B_i)$$

however

$$\sum_{\{i:B_i \text{ small}\}} \frac{|B_i|}{|B|}\,\delta(A|B_i) \leq \frac{N/M\ \delta^3 M/20}{\delta N} \leq \frac{\delta^2}{20}\,N$$

Hence

$$\delta + \frac{\delta^2}{20} \leq \sum_{\{i:B_i \text{ large}\}} \frac{|B_i|}{|B|}\delta(A|B_i) \leq \max_{\{i:B_i \text{ large}\}} \delta(A|B_i)$$

and this is what the proposition states. □

At this point, it is enough to cover the set A, with a union of disjoint progressions $B_1, \ldots, B_k$ such that the number of them is not more then: $k \leq \frac{N}{\log\log(N)}$, and moreover their union B has at most $(1-\delta/10)\,N$ elements. Indeed, then by the above proposition, one of them, say B_i, has the property that: $|B_i| \geq \delta^3/20\ \log\log(N)$ and $\delta(A|B_i) \geq (\delta + \delta^2/20)$. Then we achieved that the density of A increased in a "long" arithmetic progression.

We have to establish a few simple facts about partitioning sets into progressions of a common difference d, whose proof is again left as an exercise.

Definition 2. *Let $A \subset \mathbb{N}$ be a finite set, and let d be a natural number. We say that $a \in A$ and $b \in A$ are equivalent, and write $a \sim b$ if there is a progression:*

$$P = \{a, a+d, \ldots, a+sd = b\} \subset A \quad \text{or} \quad P = \{b, b+d, \ldots, b+sd = a\} \subset A$$

Proposition 3. *One has*

- (i) *The relation "$\sim$" is an equivalence relation. The equivalence classes are maximal progressions: $A = P_1 \cup \ldots \cup P_k$ of common difference d.*
- (ii) *One has: $k = |(A+d)\backslash A|$.*
- (i3) *The complement of A is partitioned into at most $k+d$ progressions, that is:*

$$A^c = B_1 \cup \ldots \cup B_l \quad \text{where} \quad l \leq k+d$$

The last key ingredient of the proof is based on the notion

Definition 3. *Let $a, d_1, \ldots, d_k$ be natural numbers. A k- cube is a set of the form*

$$M(a, d_1, \ldots, d_k) = \{a + \varepsilon_1 d_1 + \ldots + \varepsilon_k d_k :\ \varepsilon_i =\ 0 \text{ or } 1\ \forall\ 1 \leq i \leq k\}$$

Note, that a 1- cube is just a pair of points: $M_1 = \{a, a+d_1\}$ a 2- cube is of the form: $M_2 = \{a, a+d_1, a+d_2, a+d_1+d_2\}$. In general a k- cube has 2^k elements. If $M_i = M(a, d_1, \ldots, d_i)$ then $M_{i+1} = M_i \cup (M_i + d_{i+1})$.

Lemma 1. *Let $0 < \delta < 1$ and k, N be natural numbers such that: $N > (2/\delta)^{2^k}$. If $A \subset [1, N]$ and $|A| = \delta N$ then a contains a k- cube.*

Proof. The number of pairs $a < b$, $a \in A$, $b \in A$ is: $\delta N(\delta N - 1)/2 \geq \delta^2/2\ N(N-1)$. Since the differences $b - a$ can take at most $N - 1$ possible values, the must be a difference, say d_1 which is the common difference of at least $k \geq \delta^2/2\ N$ pairs. That is $d_1 = b_1 - a_1 = \ldots = b_k - a_k$.

Let $A_1 = \{a_1, \ldots, a_k\}$ and $\delta_1 = \delta^2/2$, then

$$|A_1| \geq \delta_1 N \quad \text{and} \quad A_1 \cup (A_1 + d_1) \subset A$$

This establishes the lemma for $k = 1$ as $\delta^2/2\ N \geq 1$ by our assumption on N. We proceed by induction, assume the statement is true for $k - 1$. Note that

$$(2/\delta)^{2^k} = (2/\delta_1)^{2(k-1)}$$

hence A_1 contains a $(k-1)$- cube $M_{k-1} = M(a, d_2, \ldots, d_k)$. It follows that A contains the k- cube: $M(a, d_1, d_2, \ldots, d_k) = M_{k-1} \cup (M_{k-1} + d_1$. This proves the lemma. □

Corollary 1. *Let $0 < \delta < 1$ and let N be such that: $N > 2^{\frac{4}{\delta^2}}$. If $A \subset [1, N]$ and $|A| = \delta N$ then a contains a k- cube, where $k \geq \frac{1}{2}\ \log\log(N)$.*

Proof. By the above lemma, A contains a k- cube where k is the largest integer such that: $N > (2/\delta)^{2^k}$, that is when $k = [\log\log(N) - \log\log(2/\delta)]$. Here $[x]$ denotes the integer part of x.

If $N > 2^{\frac{2}{\delta^2}}$ then $\frac{1}{2}\log\log(N) > \log\log(2/\delta)$ thus $k \geq \frac{1}{2}\log\log(N)$. □

After these preparations, we turn to Szemeredi's proof of Roth' theorem.

Lemma 2. *Let Let $0 < \delta < 1$ and let N be such that: $N > 2^{\frac{4}{\delta^2}}$. If $A \subset [1, N]$ and $|A| = \delta N$ then one of the following holds.*

(i) *A contains a 3-progression*
(ii) *There is a progression P, such that $|P| \geq \delta^3/20\ \log\log(N)$ and $\delta(A|P) \geq (\delta + \delta^2/20)$*

Proof. Assume A does not contain a 3-progression. For $i = 0, 1, 2, 3$ let $A_i = A \cap [iN/4, (i+1)N/4]$. If $|A_i| \leq \frac{D}{2}\,N/4$ for some i, then $A_j \geq (\delta + \frac{D}{6})\,N/4$ for some j. But the interval $[iN/4, (i+1)N/4]$ is a progression on which the density of A is increased to $\delta + \frac{D}{6}$, and case (ii) is satisfied. So we assume now that $|A_i|\frac{D}{8}\,N$ for all i.

Decompose $[N/4, N/2]$ into intervals of length $\sqrt{N}$, one of them, say I has the property that $\delta(A|I) \geq \delta/2$. By corollary 1, $A \cap I$ contains a cube of dimension

$$k \geq \frac{1}{2}\ \log\log(N)$$

Indeed $\log\log sqrtN) = \log\log(N) - 1$, and $|A \cap I| \geq \delta/2\,|I|$. Let us denote this cube $M(a, d_1, \ldots, d_k)$, and for $1 \leq i \leq k$ let $M_i = M(a, d_1, \ldots, d_i)$. Introduce the sets:

$$Q_i = \{2b - c : \ b \in M_i,\ c \in A_0\}$$

Then one has

$$(i) \quad Q_i \cap A = \emptyset$$

$$(ii) \quad |Q_i| \geq |A_0| \geq \frac{\delta}{8}\, N$$

$$(i3) \quad |Q_i \cup (Q_i + 2d_{i+1}) \subset Q_{i+1}$$

Indeed the points $c, b, 2b - c$ form a 3-progression, and if both c and b are in A, then by our assumption, $2b - c$ is not in A, this shows (i). For a fixed $b \in M_i$, one gets $|A_0|$ distinct points in Q_i. Finally if $x = 2b - c \in Q_i$, then $x + 2d_{i+1} = 2(b + d_{i+1}) - c \in Q_{i+1}$ since $b + d_{i+1} \in M_{i+1}$.

Now, the sets $Q_{i+1} \backslash Q_i$ are disjoint, hence one of them have size:

$$|Q_{i+1} \backslash Q_i| \leq \frac{N}{k} \leq \frac{2N}{\log\log(N)}$$

Let $B = Q_i^c$, that is the complement of Q_i, and let $d = 2d_{i+1}$. Partition B into progressions of common difference d, as in definition 2: $B = B_1 \cup \ldots \cup B_l$. By proposition 3., the number of parts is at most

$$l \leq d + |Q_{i+1} \backslash Q_i| \leq 2\sqrt{N} + \frac{2N}{\log\log(N)} \leq \frac{3N}{\log\log(N)}$$

Indeed the number of parts in B is at most d more, then the number of parts in its complement Q_i, which is bounded by $|Q_{i+1} \backslash Q_i|$, using (i3).

We are in position to apply proposition 2. Indeed $A \subset B$ by (i), moreover $|B| = N - |Q_i| \leq N(1 - \delta/8)$. It follows that there is a part B_i for which:

$$|B_i| \geq \frac{\delta^3}{60}\ \log\log(N) \quad \text{and} \quad \delta(A|B_i) \geq (\delta + \delta^2/20).$$

Choosing $P = B_i$ the conditions of case (ii) are satisfied, and the lemma is proved. $\square$

Finally one can do the induction argument as in the Fourier analytic proof.