Roth's Theorem on 3-term Arithmetic Progressions

Mustazee Rahman

1 Introduction

This article is a discussion about the proof of a classical theorem of Roth's regarding the existence of three term arithmetic progressions in certain subsets of the integers. Before beginning with this task, however, we will take a brief look at the history and motivation behind Roth's theorem. The questions and ideas surrounding this subject may have begun with a wonderful theorem due to van der Warden.

Theorem 1.1 (van der Warden, 1927)**.** *Given positive integers k and r, there exists a positive integer $N = N(r,k)$ such that if the integers in $[N]$[1] are coloured using r colours then it will contain a monochromatic non trivial* [2] *k-term arithmetic progression.*

It is a fruitful exercise for the reader to show the equivalence of van der Warden's theorem with the following statement: if all of the positive integers are coloured using r colours then there will exist monochromatic arithmetic progressions of any length. The proof given by van der Warden was an ingenious inductive argument that did not provide good bounds on the size of $N(r,k)$. Following van der Warden, Erdos and Turan proposed a stronger version of the theorem hoping that its proof would lead to a better version of van der Warden's theorem.

Conjecture 1.2 (Erdős-Turán)**.** *Given δ and k, there exists a positive integer $N = N(\delta,k)$ such that any subset $A \subset [N]$ with size $|A| \geq \delta N$ will contain a non trivial k-term arithmetic progression.*

It took a while for this conjecture to be settled in its entirety. In 1953, Roth made a contribution by proving the result for $k = 3$, which is the subject of this article.

Theorem 1.3 (Roth, 1953)**.** *For $\delta > 0$ and $N > \exp(\exp(C\delta^{-1}))$ with some absolute constant C, if $A \subset [N]$ has size $|A| = \delta N$ then A contains a non trivial arithmetic progression of length 3.*

The parameter δ is referred to as the **density** of A within $[N]$. Roth's proof makes use of Fourier analysis. We will give a streamlined version due to Gowers. To begin, we will 'embed' the problem in $\mathbb{Z}/N\mathbb{Z}$ by considering 3-term arithmetic progressions (which we shall abbreviate as 3-AP) of A modulo N. Despite losing data about A in this manner, we will have the advantage of working in the group $\mathbb{Z}/N\mathbb{Z}$ where we can apply Fourier analysis. The proof has essentially two parts: if the set A is 'random' then we can exhibit lots of 3-APs, while if A has structure then we can find a sub-progression of $[N]$ and a subset of A which has larger density within the sub-progression. We then iterate the procedure until we get a subset of A with high enough density within some sub-progression of $[N]$. Once the density is high enough, it is easy to exhibit non trivial 3-APs.

Roth's argument becomes hard to generalize to 4-APs or longer, but it was recently extended with the work of Gowers. Returning to the Erdős-Turán conjecture, in 1969 Szemerédi gave a proof for $k = 4$ and later extended his proof for any k. His proof was a tour-de-force of ingenious and

[1] $[N] = \{1, 2, \ldots, \mathrm{N}\}$ for any positive integer N.

[2] Monochromatic means that all terms of the progression have the same colour and non trivial means that the constant difference of the terms is not zero.

sophisticated combinatorics, but it still did not establish any good bounds on the van der Warden numbers. A different approach was opened up by Furstenberg in 1977 who used ergodic theoretic methods to prove the conjecture, which by that time was deservedly known as Szemerédi's theorem. Despite not giving any good bounds of $N(\delta, k)$, the ergodic methods did allow for generalizations that were previously inaccessible. For example, it led to a multi-dimentional version of Szemerédi's theorem and allowed for the constant difference of the APs to have special forms (e.g. squares). It was not until 2001 when Gowers extended Roth's Fourier analytic methods that eventually lead to good bounds of the van der Warden numbers.

Theorem 1.4 (Gowers, 2001)**.** *There exists a positive constant c_k and an absolute constant C such that any $A \subset [N]$ with size $|A| > CN/(\log\log N)^{c_k}$ contains a non trivial k-term arithmetic progression.*

One of Gowers's insights was the development of "quadratic Fourier Analysis" which replaced Roth's "linear Fourier Analysis" in the proof of Szemerédi's theorem. Gowers's work has transformed the subject and lead to other beautiful results. Most notable is perhaps the work of Green and Tao on arithmetic progressions within primes.

Theorem 1.5 (Green & Tao, 2003)**.** *The primes contain arbitrarily long non trivial arithmetic progressions.*

2 Proof of Roth's Theorem

The Fourier transform of a function $f : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ is given by

$$\hat{f}(r) = \sum_{k=0}^{N-1} f(k) e^{-\frac{2\pi i}{N} rk} \tag{2.1}$$

For $f,\ g : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$, the Fourier transform of $x \mapsto \sum_k f(k)\overline{g(k-x)}$ is $\hat{f}(r)\overline{\hat{g}(r)}$. Thus

$$|\hat{f}(r)||\hat{g}(r)| \leq \sum_x |\ \sum_k f(k)\overline{g(k-x)}\ | \tag{2.2}$$

This inequality is important to us since it shows that if both f and g have a large common Fourier coefficient (say to the order of N) then f must have a large inner-product with some translate of g. Another important identity is Plancherel's formula

$$\sum_{k=0}^{N-1} |f(k)|^2 = \frac{1}{N} \sum_{k=0}^{N-1} |\hat{f}(k)|^2 \tag{2.3}$$

Finally there are the orthogonality relations of the functions $x \mapsto e^{\frac{2\pi i}{N} kx}$ which gives

$$\frac{1}{N} \sum_{k=0}^{N-1} e^{\frac{2\pi i}{N} kx} = \begin{cases} 1 & \text{if } x \equiv 0 \mod N \\ 0 & \text{otherwise} \end{cases} \tag{2.4}$$

Setup: We begin with $A \subset \{0, 1, \ldots, N-1\}$ of size $|A| = \delta N$ and identify it with the *set* $\mathbb{Z}/N\mathbb{Z}$ in the natural way. A triple of natural numbers x, y, z forms a 3-AP if and only if $x + z = 2y$. As such we will first count triples satisfying this equation in $\mathbb{Z}/N\mathbb{Z}$ with $x, y, z \in A$ with using the Fourier transform. Let S_0 be the number of triples $x, y, z \in A$ solving $x + z = 2y$ in $\mathbb{Z}/N\mathbb{Z}$, and let $\mathbf{1}_A$ denote the indicator function of A. The Fourier transform of the indicator satisfies $\hat{\mathbf{1}}_A(0) = |A| = \delta N$. Using the orthogonality relations (2.4) we get that

$$S_0 = \sum_{x,y,z \in A} \frac{1}{N} \sum_{k=0}^{N-1} e^{-\frac{2\pi i}{N}(x+z-2y)k} = \frac{1}{N} \sum_{k=0}^{N-1} \widehat{\mathbf{1}_A}(k)^2 \widehat{\mathbf{1}_A}(-2k) \tag{2.5}$$

$$= \delta^3 N^2 + \frac{1}{N} \sum_{k=1}^{N-1} \widehat{\mathbf{1}_A}(k)^2 \widehat{\mathbf{1}_A}(-2k) \tag{2.6}$$

The leading term $\delta^3 N^2$ in (2.6) is suggestive since it would be the number of triples in A satisfying $x + z = 2y$ if the elements in A were picked independently from $\{0, \ldots, N-1\}$ each with probability δ. An attempt to estimate how close S_0 is to this leading term leads to a definition.

Definition 2.1. A set $A \subset \{0, \ldots, N-1\}$ is ϵ-uniform if $|\widehat{\mathbf{1}_A}(k)| \leq \epsilon N$ for all $k \neq 0$.

Notice that 3-APs in A over $\mathbb{Z}/N\mathbb{Z}$ are not always 3-APs in A over $\mathbb{Z}$. Hence we denote by S the number of triples $x, y, z \in A$ such that $x + z = 2y$. The key observation to make is that if $x, y \in M_A = A \cap [N/3, 2N/3)$ then a 3-AP in A over $\mathbb{Z}/N\mathbb{Z}$ (we shall call these $\mathbb{Z}/N\mathbb{Z}$-progressions) is also a 3-AP in A over $\mathbb{Z}$ (which we shall call $\mathbb{Z}$-progressions).

Lemma 2.2. *If A is ϵ-uniform for $\epsilon \leq \frac{\delta^2}{8}$ and $|M_A| \geq \frac{\delta N}{4}$ then $S \geq \frac{\delta^3 N^2}{32}$.*

Proof. Computing as in (2.6) and counting solutions to $x + z = 2y$ with $x, y \in M_A$ we get

$$S \geq \frac{1}{N} \sum_{k=0}^{N-1} \widehat{\mathbf{1}_{M_A}}(k) \widehat{\mathbf{1}_A}(k) \widehat{\mathbf{1}_{M_A}}(-2k) = \delta |M_A|^2 + \frac{1}{N} \sum_{k=1}^{N-1} \widehat{\mathbf{1}_{M_A}}(k) \widehat{\mathbf{1}_A}(k) \widehat{\mathbf{1}_{M_A}}(-2k)$$

By Cauchy-Schwarz and Plancherel's formula,

$$\begin{aligned}
|\sum_{k=1}^{N-1} \widehat{\mathbf{1}_{M_A}}(k) \widehat{\mathbf{1}_A}(k) \widehat{\mathbf{1}_{M_A}}(-2k)| &\leq \epsilon N \sum_{k=1}^{N-1} |\widehat{\mathbf{1}_{M_A}}(k) \widehat{\mathbf{1}_{M_A}}(-2k)| \\
&\leq \epsilon N (\sum_{k=1}^{N-1} |\widehat{\mathbf{1}_{M_A}}(k)|^2)^{\frac{1}{2}} (\sum_{k=1}^{N-1} |\widehat{\mathbf{1}_{M_A}}(-2k)|^2)^{\frac{1}{2}} \\
&\leq \epsilon N \sum_{k=0}^{N-1} |\widehat{\mathbf{1}_{M_A}}(k)|^2 \\
&= \epsilon N^2 \sum_{k=0}^{N-1} |\mathbf{1}_{M_A}(k)|^2 \\
&= \epsilon N^2 |M_A|
\end{aligned}$$

The fact that $\epsilon \leq \frac{\delta^2}{8}$ and $|M_A| \geq \frac{\delta N}{4}$ implies

$$S \geq \delta |M_A|^2 - \epsilon N |M_A| \geq \frac{1}{2} \delta |M_A|^2 \geq \frac{1}{32} \delta^3 N^2$$

□

There are $|A|$ trivial 3-APs $x = y = z$. Thus Lemma 2.2 guarantees the existence of a non trivial 3-AP provided that $N > 32\delta^{-2}$. To summarize, if $A \subset \{0, \ldots, N-1\}$ contains no non-trivial 3-APs then one of the following conditions must hold

1. $N \leq 32\delta^{-2}$

2. A is not ϵ-uniform for any $\epsilon \leq \frac{\delta^2}{8}$

3. There exists an $\mathbb{Z}$-progression P with length $|P| \geq \frac{N}{3}$ such that

$$|A \cap P| \geq (\delta + \frac{\delta}{8})|P|$$

The last condition follows from observing that if $|M_A| < \frac{\delta N}{4}$ then

$$\max\{|A \cap [0, N/3)|, |A \cap [2N/3, N)|\} \geq \frac{9}{8}\delta \cdot (\frac{N}{3})$$

Condition 3 is known as the density incrementing property and it is at the heart of the proof. We will show that if A is not ϵ-uniform as in condition 2 then some subset of A must have larger density in some sub-progression as in condition 3. To this end, we say a $\mathbb{Z}/N\mathbb{Z}$-progression P is **non-overlapping** if its length $|P|$ and its constant difference d satisfies $|P|d < N$. The point is that we can write a non-overlapping $\mathbb{Z}/N\mathbb{Z}$-progression as a disjoint union of two $\mathbb{Z}$-progressions since it avoids wrap around issues. In order to address sets that are not ϵ-uniform, we need to consider functions of mean value 0. Define the balanced indicator of A as $f_A(x) = \mathbf{1}_A(x) - \delta$, which is translated to have mean value 0. Note that $\widehat{f_A}(r) = \widehat{\mathbf{1}_A}(r)$ for $r \neq 0$.

Lemma 2.3. *If $|\widehat{\mathbf{1}_A}(r)| \geq \epsilon N$ for $r \neq 0$ then there exists a non-overlapping $\mathbb{Z}/N\mathbb{Z}$-progression B with $|B| \geq \frac{\sqrt{N}}{4}$ such that $|A \cap B| \geq (\delta + \frac{\epsilon}{4})|B|$.*

Proof. Claim: There exists a non-overlapping $\mathbb{Z}/N\mathbb{Z}$-progression B' with $|B'| \geq \frac{\sqrt{N}}{4}$ such that $|\widehat{\mathbf{1}_B}(r)| \geq \frac{1}{2}|B'|$.

Proof of Claim: Consider the N pairs $(0,0), (1,r), \ldots, ((N-1), (N-1)r)$ modulo N. They lie in $[0, N-1]^2$ which we partition into $\lfloor\sqrt{N}\rfloor^2 < N$ equal squares of side-length $l = N/\lfloor\sqrt{N}\rfloor$. Two of the pairs must lie in the same square which means that for integers $0 \leq a < b \leq N-1$ we have $b - a \leq l$ and $r(b-a) \leq l \pmod N$. Set $d = b - a$ and let B' be the non-overlapping $\mathbb{Z}/N\mathbb{Z}$-progression of length $\lfloor\frac{\lfloor\sqrt{N}\rfloor}{\pi}\rfloor$ given as $\{\ldots, -2d, -d, 0, d, 2d, \ldots\}$. We get that

$$\begin{aligned}
|\widehat{\mathbf{1}'_B}(r) - |B'|| &= |\sum_x \mathbf{1}'_B(x)(e^{-\frac{2\pi i}{N}rx} - 1)| \\
&\leq \sum_{|x| \leq \frac{1}{2}|B'|} |e^{-\frac{2\pi i}{N}rdx} - 1| \\
&\leq \frac{1}{2}|B'|\,(\frac{2\pi|B'|l}{2N}) \leq \frac{1}{2}|B'|
\end{aligned}$$

Hence $|\widehat{\mathbf{1}_B}(r)| \geq \frac{1}{2}|B'|$. The desired $\mathbb{Z}/N\mathbb{Z}$-progression of the lemma will be a translate of the progression B'. Writing $B = B' + x$, note that $|A \cap B| \geq (\delta + \frac{1}{4}\epsilon)|B| \Leftrightarrow \sum_k f_A(k)\mathbf{1}_{B'}(k-x) \geq \frac{1}{4}\epsilon|B'|$. Set $G(x) = \sum_k f_A(k)\mathbf{1}_{B'}(k-x)$. Recalling back to (2.2) and using the claim above, it follows that

$$\sum_x |G(x)| \geq |\widehat{G}(r)| \geq \frac{1}{2}\epsilon N|B'|$$

Since $G(x)$ also has mean value 0, we get that $\sum_x G(x) + |G(x)| \geq \frac{1}{2}\epsilon N|B'|$, which implies that for some x we have $G(x) + |G(x)| \geq \frac{1}{2}\epsilon|B'|$. This in turn implies that $G(x) \geq \frac{1}{4}\epsilon|B'|$ as required. $\square$

If B is a non-overlapping $\mathbb{Z}/N\mathbb{Z}$-progression on which A has density $\delta + \epsilon'$ then there exists a $\mathbb{Z}$-progression P with $|P| \geq \frac{1}{2}\epsilon'|B|$ such that A has density at least $\delta + \frac{1}{2}\epsilon'$ on P. Indeed, write $B = P_1 \sqcup P_2$ where P_1, P_2 are $\mathbb{Z}$-progressions satisfying $|P_1| \leq |P_2|$. If $|P_1| \leq \frac{1}{2}\epsilon'|B|$ then $|A \cap P_2| \geq (\delta + \epsilon')|B| - |P_1| \geq (\delta + \frac{1}{2}\epsilon')|B| \geq (\delta + \frac{1}{2}\epsilon')|P_2|$. If this is not the case then both $|P_1|$ and $|P_2|$ are at least $\frac{1}{2}\epsilon'|B|$. Thus A must have density at least $\delta + \epsilon'$ on one of them.

Combining this observation and Lemma 2.3 we deduce that if A is not ϵ-uniform then there exists a $\mathbb{Z}$-progression P with $|P| \geq \frac{1}{32}\epsilon\sqrt{N}$ and $|A \cap P| \geq (\delta + \frac{1}{8}\epsilon)|P|$. Combining this with conditions 1, 2 and 3 from before gives the final proposition.

Proposition 2.4. *Let $A \subset \{0, \ldots, N-1\}$ with $|A| = \delta N$ and $N > 32\delta^{-2}$. Then either A contains a non-trivial 3-AP or there exists an arithmetic progression $P \subset \{0, \ldots, N-1\}$ with $|P| \geq \frac{1}{256}\delta^2\sqrt{N}$ and $|A \cap P| \geq (\delta + \frac{1}{64}\delta^2)|P|$.*

Final step of the proof: Roth's theorem follows by iterating Proposition 2.4. Assume, for contradiction, that A has no non-trivial 3-APs. We will get a contradiction for large N. Applying Proposition 2.4, we get a progression P_1 of length N_1 and a subset $A_1 = P_1 \cap A$ of A such that $|A_1| = \delta_1 N_1$ where $N_1 \geq \frac{1}{256}\delta^2\sqrt{N}$ and $\delta_1 \geq \delta + \frac{1}{64}\delta^2$. Since A_1 has no non-trivial 3-AP, we iterate. After k steps we get $A_k \subset A$ contained within a progression P_k of length N_k such that A_k has no non-trivial 3-APs, $|A_k| = \delta_k N_k$ with $N_k \geq \frac{1}{256}\delta_{k-1}^2\sqrt{N_{k-1}}$ and $\delta_k \geq \delta_{k-1} + \frac{1}{64}\delta_{k-1}^2 \geq \delta + \frac{k}{64}\delta^2$. After $k = \frac{64}{\delta}$ steps, A_k has density at least 2δ within some subprogression. Repeating the procedure with this new set an additional $k = \frac{64}{2\delta}$ times, we get a subset A_k' of density at least 4δ within some subprogression. In general after $k = \frac{64}{\delta}(1 + \ldots + \frac{1}{2^{l-1}})$ steps the set A_k has density at least $2^l\delta$ within some sub-progression. The upshot is that after $\frac{128}{\delta}$ steps, the density is guaranteed to exceed 1 (in fact it is infinite). Since no set can have density exceeding 1 within a non-empty progression, it suffices to find a bound on N which guarantees a non-empty sub-progression even after $\frac{128}{\delta}$ steps.

Using the fact that $N_k \geq \frac{1}{256}\delta_{k-1}^2\sqrt{N_{k-1}}$, we deduce that after k steps the sub-progression in our hands will have length at least $\frac{1}{256^2}\delta^4 N^{\frac{1}{2^k}}$. Therefore, we simply need to show that for $k = \frac{128}{\delta}$, $N^{\frac{1}{2^k}} \geq 256^2\delta^{-4} = 2^{16}\delta^{-4}$. After taking logarithms and substituting for k, this reduces to

$$\log(\mathrm{N}) \geq [16\log(2) + 4\log(\delta^{-1})]2^{128\delta^{-1}}$$

It is an easy computation to verify that for $0 < \delta \leq 1$, $16\log(2) + 4\log(\delta^{-1}) \leq 2^{4\delta^{-1}}$. Hence for the desired contradiction it suffices to have

$$N \geq \exp(\exp(132\log(2) \cdot \delta^{-1}))$$

This establishes Roth's theorem with an absolute constant $C < 100$.

3 Example of large set with no 3-AP

We now take up the task of showing that there are indeed large subsets of $[N]$ with no 3-APs. Going back to the proof of Roth's theorem, we saw that if $A \subset [N]$ of size δN was picked randomly such that each element belonged to A independently with probability δ then we expected $\delta^3 N^2 = |A|^3/N$ 3-APs in A. Since there are $|A|$ trivial such 3-APs, one might expect that if $|A| > N^{\frac{1}{2}}$ then A will contain a non-trivial 3-AP. Yet Behrend produced an example of a surprisingly large subset of $[N]$ with no 3-APs.

Theorem 3.1 (Behrend, 1946)**.** *There exists $A \subset [N]$ with no 3-term arithmetic progressions but $|A| > N\exp(-c\sqrt{\log N})$ with c an absolute constant.*

Proof. The construction rests on the seemingly innocent fact that a line intersects a sphere in at most two points. Consider n-tuples of integers $(x_1, \ldots, x_n)$ from $[0, d]^n$. There are $(d+1)^n$ such tuples while the function $(x_1, \ldots, x_n) \mapsto \sum_i x_i^2$ takes on at most nd^2 values. Hence for some integer r, there must be at least $\frac{(d+1)^n}{nd^2}$ n-tuples that satisfy $r = \sum_i x_i^2$. In other words, the sphere S_r of radius $\sqrt{r}$ contains plenty of lattice points.

Now take the set $A = \{\sum_{i=1}^n x_i(2d+1)^{i-1}\}$ where each $(x_1, \ldots, x_n)$ is an integer tuple lying on S_r as above. By size considerations, it is easy to see that the elements of A are all distinct. For example, if $\sum_{i=1}^n (x_i - y_i)(2d+1)^{i-1} = 0$ then $x_n - y_n = 0$, for otherwise if $x_n - y_n > 0$ then $\sum_{i=1}^n (x_i - y_i)(2d+1)^{i-1} \geq (2d+1)^{n-1} - [(2d+1)^{n-2} - 1] > 0$. An analogous argument with $x_n - y_n < 0$ gives another contradiction. Inductively it follows that each $x_i = y_i$. Thus $|A| \geq \frac{(d+1)^n}{nd^2}$. Notice too that the elements of A are all less than $(2d+1)^n$.

If A contained a 3-AP then

$$\sum_i x_i(2d+1)^{i-1} + \sum_i z_i(2d+1)^{i-1} = \sum_i 2y_i(2d+1)^{i-1}$$

would imply that $x_i + z_i = 2y_i$ by size considerations again. Thus $(y_1, \ldots, y_n)$ would be the midpoint of $(x_1, \ldots, x_n)$ and $(z_1, \ldots, z_n)$ while all three of these points lie on a common sphere. This is impossible. Finally take n to be about $\sqrt{\log(N)}$ and d to be about $e^{\sqrt{\log(N)}}$ to get a set $A \subset [N]$ of size $|A| \geq N\exp(-c\sqrt{\log(N)})$ with no 3-APs. □

References

[1] T. Gowers, *Additive and Combinatorial Number Theory*, 2007.
`http://www.dpmms.cam.ac.uk/~wtg10/addnoth.notes.ps`

[2] N. Lyall, *Roth's Theorem: The Fourier Analytic approach*, 2005.
`http://www.math.uga.edu/%7Elyall/REU/Roth.pdf`

[3] K. Soundararajan, *Lecture Notes on Additive Combinatorics: Winter 2007*, 2007.
`http://math.stanford.edu/~ksound/Notes.pdf`