ROTH'S THEOREM
THE FOURIER ANALYTIC APPROACH

NEIL LYALL

Roth's Theorem. *Let $\delta > 0$ and $N \geq \exp\exp(C\delta^{-1})$ for some absolute constant C. Then any $A \subset \{0, 1, \ldots, N-1\}$ of size $|A| = \delta N$ necessarily contains a (non-trivial) arithmetic progression of length three.*

1. Embedding the problem in $\mathbb{Z}_N$ and the Fourier transform

If x, y, z are natural numbers and $x + y = 2z$ then they form an arithmetic progression of length three, which we'll call briefly a 3-progression. Instead of counting these directly we shall instead first count the number of triples x, y, z chosen from the set A which satisfy the same equation in $\mathbb{Z}_N$, that is modulo N. Although in doing this we lose some information about the arithmetic progessions in A, the up side is that we have now embedded the problem in a group on which we can do Fourier analysis.

Recall that the (discrete) Fourier transform $\hat{f}$, of a function $f : \mathbb{Z}_N \to \mathbb{C}$ is defined by

$$\hat{f}(k) = \sum_{x=0}^{N-1} f(x) e^{-\frac{2\pi i}{N} xk}. \tag{1}$$

It is clear from the definition that $\hat{f}(0) = \sum_x f(x)$ and that for any k the Fourier coefficients satisfy the inequality

$$|\hat{f}(k)| \leq \sum_x |f(x)|. \tag{2}$$

Using this and the fact that the Fourier transform of $\sum_y f(y)\overline{g(y-x)}$ is $\hat{f}(k)\overline{\hat{g}(k)}$ we see that

$$|\hat{f}(k)||\hat{g}(k)| \leq \sum_x \Big| \sum_y f(y)\overline{g(y-x)} \Big|. \tag{3}$$

This inequality will be of crucial importance to us, it allows us to conclude that if two functions f and g have at least one large Fourier coefficient in common, by which we mean they have size proportional to N, then f must have a large inner product with at least one translate of g.

We shall also make use of Plancherel's identity

$$\sum_x |f(x)|^2 = \frac{1}{N} \sum_k |\hat{f}(k)|^2, \tag{4}$$

which is, at least in this setting, an immediate consequence of the orthogonality relation

$$\frac{1}{N} \sum_{k=0}^{N-1} e^{\frac{2\pi i}{N} xk} = \begin{cases} 1 & x \equiv 0 \mod N \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

2. Outline of the Proof

We begin with $A \subset \{0, 1, \dots, N-1\}$ with $|A| = \delta N$. We will take the liberty of assuming that N is odd, if N is even we can always consider A instead as a subset of $\{0, 1, \dots, N\}$, this has the effect of only changing the density very slightly[1]. Our stratagy could be roughly summarized as follows;

(i) If A is "suitably random", then we shall be able to conclude that A contains many arithmetic progressions of length three.
(ii) If A is not "suitably random", then we shall show that there exists a long arithmetic progression P such that A has increased density on P, that is

$$|A \cap P| \geq (\delta + \varepsilon')|P|,$$

for some $\varepsilon' > 0$ that depends only on δ.

So if A is not "suitably random" then we turn our attention to a subset $A_1 \subset A$ (namely $A \cap P$) which has higher density in some sub-progression of the integers. We must then determine if A_1 is "suitably random" or not, we iterate this argument until the theorem is proven. We shall show that if A contains no 3-progressions then there must be a progression on which a subset of A has density greater than one, which is absurd. For this argument to work, we must count the number of necessary iterations to reach this density and ensure that we have not eliminated too many elements in the process (in order to arrive at our contradiction we must ensure that our final progression is non-empty).

3. A suitable notion of randomness

With the aid of the Fourier transform we can now count the number of solutions to the congruence

$$x + y \equiv 2z \mod N,$$

with $x, y, z \in A$. Let $\mathcal{N}_0$ denote the number of triples solving the above congruence, then by the orthogonality relation (5) and the definition of the Fourier transform one has

$$\mathcal{N}_0 = \sum_{x\in A}\sum_{y\in A}\sum_{z\in A}\frac{1}{N}\sum_{k=0}^{N-1} e^{-\frac{2\pi i}{N}(x+y-2z)k} = \frac{1}{N}\sum_{k=0}^{N-1}\widehat{1_A}(k)^2\widehat{1_A}(-2k),$$

where $1_A(x)$ denotes the characteristic function of the set A. Using the fact that $\widehat{1_A}(0) = |A| = \delta N$ we see that

$$\mathcal{N}_0 = \delta^3 N^2 + \frac{1}{N}\sum_{k=1}^{N-1}\widehat{1_A}(k)^2\widehat{1_A}(-2k).$$

The leading term $\delta^3 N^2$ above is instructive as it is the number of solutions of the congruence if the set A were random, obtained by selecting each natural number from 1 to N independently with probability δ. Indeed after choosing say x and z arbitrarily from the set A, which can be done $\delta^2 N^2$ ways, the probability that $y \equiv 2z - x \mod N$ is in A is equal to δ. Thus if A is very uniformly distributed among the numbers $\{0, 1, \dots, N-1\}$ then we expect it to contain a lot of 3-progressions mod N.

[1] The new density δ' satisfies $(1 - \frac{1}{N})\delta < \delta' < \delta$.

Notice that if $|\widehat{1_A}(k)| \leq \varepsilon N$ for all $k \neq 0$, then it would follow from Plancherel's identity that

$$\Big|\sum_{k=1}^{N-1} \widehat{1_A}(k)^2 \widehat{1_A}(-2k)\Big| \leq \max_{k\neq 0} |\widehat{1_A}(-2k)| \sum_{k=0}^{N-1} |\widehat{1_A}(k)|^2 \leq \varepsilon N^2 \sum_{x=0}^{N-1} |1_A(x)|^2 = \varepsilon\delta N^3,$$

and hence that

$$\mathcal{N}_0 \geq \delta^3 N^2 - \varepsilon\delta N^2. \tag{6}$$

We therefore see that the parameter ε that we introduced is, in the particular sense above, measuring how close A is to being random. With this in mind we make the following definition.

Definition 1. We say that A is **ε-uniform** (or **suitably random**) if

$$|\widehat{1_A}(k)| \leq \varepsilon N,$$

for all $k = 1, \ldots, N-1$.

4. Sets which are suitably random

It is clear from the discussion above, in particular inequality (6), that if $\varepsilon < \delta^2/2$ and A is ε-uniform, then $\mathcal{N}_0 \geq \delta^3 N^2/2$, that is A contains at least $\delta^3 N^2/2$ 3-progressions mod N (including the trivial ones). We must however distinguish between arithmetic progressions mod N (which we shall refer to as $\mathbb{Z}_N$-progressions) and genuine arithmetic progressions (which we shall refer to as $\mathbb{Z}$-progressions). Let $\mathcal{N}$ denote the number of solutions in A of the equation $x + y = 2z$, that is the number of $\mathbb{Z}$-progressions of length three that are contained in A.

The key observation is that if $x, z \in M_A = A \cap [N/3, 2N/3)$, then any $\mathbb{Z}_N$-progression will also be a $\mathbb{Z}$-progression. Using this fact and a slight modification of the argument above gives the following.

Lemma 1. *If A is ε-uniform with $\varepsilon < \frac{\delta^2}{8}$ and $|M_A| \geq \frac{\delta}{4}N$, then $\mathcal{N} \geq \frac{\delta^3 N^2}{32}$.*

Proof. As before we can write

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

It follows from the Cauchy-Schwarz inequality and Plancherel's identity that

$$\begin{aligned}
\Big|\sum_{k=1}^{N-1} \widehat{1_{M_A}}(k)\widehat{1_A}(k)\widehat{1_{M_A}}(-2k)\Big| &\leq \max_{k\neq 0} |\widehat{1_A}(k)| \sum_{k=0}^{N-1} |\widehat{1_{M_A}}(k)\widehat{1_{M_A}}(-2k)| \\
&\leq \varepsilon N \Big(\sum_k |\widehat{1_{M_A}}(k)|^2\Big)^{1/2} \Big(\sum_k |\widehat{1_{M_A}}(-2k)|^2\Big)^{1/2} \\
&= \varepsilon N \sum_k |\widehat{1_{M_A}}(k)|^2 \\
&= \varepsilon N^2 \sum_x 1_{M_A}(x) \\
&\leq \varepsilon N^2 |M_A|.
\end{aligned}$$

It now follows that if we take $\varepsilon < \delta^2/8$, then

$$\mathcal{N} \geq \tfrac{1}{2}\delta|M_A|^2 \geq \delta^3 N^2/32. \qquad \square$$

So far we have not excluded the trivial 3-progressions $x = y = z$, fortunately there are only $|A| = \delta N$ of them. It is then easy to check that Lemma 1 guarantees the existence of a non-trivial 3-progression as long as one insists that $N \geq \frac{8}{\delta^2}$.

Let us summarize what we have shown so far.

Proposition 1. *Let $A \subset \{0, 1, \ldots, N-1\}$ with $|A| = \delta N$. If A contains no non-trivial $\mathbb{Z}$-progressions of length three, then one of the following must hold.*

(i) $N \leq 8\delta^{-2}$
(ii) *There exists a $\mathbb{Z}$-progression P of length $|P| \geq N/3$ such that*
$$|A \cap P| \geq (\delta + \delta/8)|P|$$
(iii) *A is not ε-uniform for any $\varepsilon \leq \delta^2/8$*

Proof. Parts (i) and (iii) follow from Lemma 1 and the discussion above. Possiblity (ii) is simply the observation that if $|M_A| < \frac{\delta}{4}N$ then
$$\max\{|A \cap [0, N/3)|, |A \cap [2N/3, N)|\} \geq 3\delta N/8 = 9\delta/8 \cdot (N/3). \qquad \square$$

5. Sets which are not suitably random

Now we arrive at the heart of the proof, where we show that a set A which is not ε-uniform necessarily has increased density on an arithmetic progression of large size.

We say that a $\mathbb{Z}_N$-progression P is *non-overlapping* if its length L and common difference d satisfy $dL < N$, the point of such a definition being that a non-overlapping $\mathbb{Z}_N$-progression is a union of two $\mathbb{Z}$-progressions.

Lemma 2. *Suppose that B' is a non-overlapping $\mathbb{Z}_N$-progression on which A has density $\delta + \varepsilon'$. Then there is a $\mathbb{Z}$-progression P of length at least $\frac{1}{2}\varepsilon'|B'|$ on which A has density at least $\delta + \frac{1}{2}\varepsilon'$.*

Proof. Write $B' = P_1 \cup P_2$, with $|P_1| \leq |P_2|$. If $|P_1| \leq \frac{1}{2}\varepsilon'|B'|$, then
$$|A \cap P_2| \geq (\delta + \varepsilon')|B'| - |P_1| \geq (\delta + \tfrac{1}{2}\varepsilon')|B'| \geq (\delta + \tfrac{1}{2}\varepsilon')|P_2|,$$
while if this is not the case then both P_1 and P_2 must have length at least $\frac{1}{2}\varepsilon'|B'|$ and consequently A must have density at least $\delta + \varepsilon'$ on one of them. $\square$

So we now know how to pass from non-overlapping $\mathbb{Z}_N$-progressions to $\mathbb{Z}$-progressions.

Lemma 3. *If $|\widehat{1_A}(r)| \geq \varepsilon N$ for some $r \neq 0$, then there exists a non-overlapping $\mathbb{Z}_N$-progression B' of length at least $\sqrt{N}/4$ such that*
$$|A \cap B'| \geq (\delta + \tfrac{1}{4}\varepsilon)|B'|.$$

For technical reasons it shall be convenient in the proof of Lemma 3 to consider functions of mean value zero. We define the *balanced* function of A to be
$$f_A(x) = 1_A(x) - \delta = \begin{cases} 1 - \delta & x \in A \\ -\delta & x \notin A. \end{cases}$$
We note that indeed $\widehat{f_A}(0) = \sum_x f_A(x) = 0$ and that $\widehat{f_A}(k) = \widehat{1_A}(k)$ for all $k \neq 0$.

Proof of Lemma 3. The key observation here is that if we write $B' = B + x$, then

$$|A \cap (B+x)| \geq (\delta + \tfrac{1}{4}\varepsilon)|B| \iff \sum_y f_A(y) 1_B(y-x) \geq \tfrac{1}{4}\varepsilon|B|.$$

Therefore our aim is now to find a long arithmetic progression B whose Fourier transform at r is also large, since by (3) this would imply that $1_A(x)$ has a large inner product with a translate of $1_B(x)$, which as we have just observed is exactly what we want.

Step 1: For any $1 \leq r \leq N-1$ there exists a non-overlapping progression B of length at least $\sqrt{N}/4$ with the property that

$$|\widehat{1_B}(r)| \geq \frac{1}{2}|B|.$$

Proof of Step 1: Fix r. If we partition $[0, N-1]^2$ into less than N (say $\lceil \sqrt{N}-1 \rceil^2$) equal squares it follows from the 'principle of the pigeons', by considering either of the collections of pairs

$$\{(0,0), (1,r), \ldots, (N-1, (N-1)r)\}$$
$$\text{or } \{(N-1,0), (N-2,r), \ldots, (0,(N-1)r)\},$$

that there exist integers $0 \leq \ell < k \leq N-1$ such that both

$$k - \ell \leq \sqrt{N} \quad \text{and} \quad r(k-\ell) \leq \sqrt{N} \mod N.$$

Let $d = k - \ell$. Define B to be the following progression of length $|B| = \lfloor \sqrt{N}/\pi \rfloor$ with common difference d;

$$\{\ldots, -2d, -d, 0, d, 2d, \ldots\}.$$

We then have

$$\begin{aligned} \big|\widehat{1_B}(r) - |B|\big| &\leq \Big| \sum_x 1_B(x)[e^{-\frac{2\pi i}{N}xr} - 1] \Big| \\ &\leq \sum_{|\ell| \leq \frac{1}{2}|B|} |e^{-\frac{2\pi i}{N}\ell d r} - 1| \\ &< \tfrac{1}{2}|B|\big(\tfrac{2\pi|B|\sqrt{N}}{2N}\big) \leq \tfrac{1}{2}|B|. \end{aligned}$$

Step 2: As advertised our $\mathbb{Z}_N$-progression B' will be a translate of the $\mathbb{Z}_N$-progression B obtained in Step 1. We must therefore now show that there is indeed a value of x for which

$$\sum_y f_A(y) 1_B(y-x) \geq \tfrac{1}{4}\varepsilon|B|.$$

Proof of Step 2: Let $G(x) = \sum_y f_A(y) 1_B(y-x)$. It follows from Step 1, the assumption that A is not ε-uniform, and (3) that

$$\sum_x |G(x)| \geq |\widehat{G}(r)| \geq \tfrac{1}{2}\varepsilon N|B|.$$

Using the fact that G has mean value zero we the see that

$$\sum_x \{|G(x)| + G(x)\} \geq \tfrac{1}{2}\varepsilon N|B|,$$

and hence for some x we must have

$$|G(x)| + G(x) \geq \tfrac{1}{2}\varepsilon|B|,$$

which in turn implies that $G(x) \geq \frac{1}{4}\varepsilon|B|$ as required.

□

Combining the two lemmata of this section we obtain

Proposition 2. *If $|\widehat{1_A}(r)| \geq \varepsilon N$ for some $r \neq 0$, then there exists a $\mathbb{Z}$-progression P of length at least $\frac{1}{32}\varepsilon\sqrt{N}$ such that*

$$|A \cap P| \geq (\delta + \tfrac{1}{8}\varepsilon)|P|.$$

6. Putting it all together and proving Roth's theorem

Combining Propositions 1 and 2 gives the following.

Proposition 3. *Let $A \subset \{0, 1, \dots, N-1\}$ with $|A| = \delta N$ and $N \geq 8\delta^{-2}$. Then either A contains a non-trivial $\mathbb{Z}$-progression of length three, or there exists a $\mathbb{Z}$-progression P of length $|P| \geq \frac{1}{256}\delta^2\sqrt{N}$ such that*

$$|A \cap P| \geq (\delta + \tfrac{1}{64}\delta^2)|P|.$$

An easy iteration argument now concludes the proof of Roth's theorem - with constant $C < 100$.

Proof of Roth's Theorem. We assume that A contains no non-trivial 3-progressions. This will, for N large enough, lead us to a contradiction.

Denote by P_1 the $\mathbb{Z}$-progression guaranteed to us by Proposition 3. We now identify $P_1 \simeq \{0, 1, \dots N_1 - 1\}$ and $A_1 \simeq A \cap P_1$ - by simply enumerating the elements of P_1 in increasing order. We know, by assumption, that A_1 contains no (non-trivial) 3-progression. In addition we know that $|A_1| = \delta_1 N_1$, where $N_1 \geq \frac{1}{256}\delta^2\sqrt{N}$ and $\delta_1 \geq (\delta + \frac{1}{64}\delta^2)$.

We now iterate this argument. In doing so $k = \frac{64}{\delta}$ times we arrive at a set $A_k \subset \{0, 1, \dots, N_k - 1\}$ with density $\delta_k \geq \delta + \delta = 2\delta$ since the density at each step is increasing by at least $\frac{\delta^2}{64}$. After another $\frac{64}{2\delta}$ steps the density is increasing from 2δ to 4δ, ...

- A density of at least $2^\ell\delta$ is reached in no more than $\frac{64}{\delta}(1 + \frac{1}{2} + \cdots + \frac{1}{2^{\ell-1}})$ steps and consequently a density exceeding one (in fact it grows to infinity) in no more than $\frac{128}{\delta}$ steps.

It is of course impossible for any subset of A to have density greater than one in a non-empty $\mathbb{Z}$-progression. We now calculate what bound must be imposed on N in order to guarantee that after this number of steps we have not eliminated all of our elements.

At each step of the iteration the size of the subprogression chosen is about the square root of the progression of the previous step. After the first step the size of the progression is at least $\frac{1}{256}\delta^2\sqrt{N}$.

- After k steps we are reduced to a progression of length at least $\frac{1}{256^2}\delta^4 N^{1/2^k}$.

It therefore remain to show that if $k = \frac{128}{\delta}$, then $N^{1/2^k} \geq 2^{16}\delta^{-4}$. Taking logs we see that this inequality is equivalent to insisting that

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

It is then an easy exercise to see that

$$16 \log 2 + 4 \log \delta^{-1} \leq 2^{4\delta^{-1}},$$

and hence that in order to arrive at a contradiction (a non-empty $\mathbb{Z}$-progression on which a subset of A has density greater than one) it is sufficient to have

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

□