A Proof of Roth's Theorem on Arithmetic Progressions

Bora Çalım

June 15, 2023

Abstract

Roth's theorem on arithmetic progressions states that if $r(N)$ denotes the cardinality of the largest subset of $\{1, 2, \dots, N\}$ which contains no nontrivial 3-term arithmetic progressions, then $\lim_{N\to\infty} \frac{r(N)}{N} = 0$. We give a proof of this theorem, following the sketch in O'Bryant's Mathoverflow answer [OBr10] and chapter 10.2 of Vaughan's book [Vau97].

1 Introduction

1.1 The problem

Let $\mathbb{N}$ be the set of positive integers (in particular, $0 \notin \mathbb{N}$). We call a subset A of $\mathbb{N}$ "good"[1] if the only solutions to $a + b = 2c$ in A are the trivial solutions, i.e., $a = b = c$. By making the change of variables $d = c - a$, we see that A is good if and only if it contains no triple of the form $(a, a + d, a + 2d)$ with $d \neq 0$.

Let $r(N)$ denote the largest size (cardinality) of a good subset of $\{1, 2, \dots, N\}$, and define $\rho(N) = \frac{r(N)}{N}$. In [Rot53], Roth proved that $\rho(N) \to 0$ as $N \to \infty$[2]. This can be phrased in words as "if a set is large, then it contains a 3-term arithmetic progression". In this article we give a proof of this result, based on the sketch given in O'Bryant's Mathoverflow answer [OBr10] and chapter 10.2 of Vaughan's book [Vau97]. The proof utilizes Fourier analysis, in the form of manipulating finite exponential sums.

Remark. In [Rot53], $\rho(N) = O(\frac{1}{\log \log N})$ is proved. This bound has been improved many times, with the most recent improvement by Kelley and Meka [KM23], showing that $\rho(N) \leq 2^{-\Omega((\log N)^c)}$ for some absolute constant c, where $\Omega(f)$ denotes a quantity that is larger in absolute value than kf for some constant k.

1.2 Notation

$O^*(f)$[3] denotes a quantity smaller in absolute than f. For instance, $f(x) = g(x) + O^*(h(x))$ means $|f(x) - g(x)| \leq h(x)$. We write $e(x)$ for $exp(2\pi i x)$. We write $[x]$ for the floor of x, i.e., the largest integer not greater than x, and $\{x\}$ for the fractional part of x, i.e., $\{x\} = x - [x]$. We write $\delta(x)$ for the function which is 1 at 0 and 0 at all other points, so that $\int_0^1 e(ax)dx = \delta(a)$. We write $1_S(x)$ for the indicator function of the set S, i.e., $1_S(x) = 1$ if $x \in S$ and $1_S(x) = 0$ if $x \notin S$.

2 Preliminaries

In this section we make several definitions and prove several lemmata which will be useful in the proof.

Lemma 1. *If $N, M \in \mathbb{N}$, then $r(N + M) \leq r(N) + r(M)$.*

Proof. Let $A \subseteq \{1, 2, \dots, N + M\}$ such that A is good and $|A| = r(N + M)$. Consider $A_1 = A \cap \{1, 2, \dots, N\}$, $A_2 = A \cap \{N + 1, N + 2, \dots, N + M\}$, so that $|A| = |A_1| + |A_2|$. Since A is good and $A_1, A_2 \subseteq A$, we see that A_1 and A_2 are good. By definition, $|A_1| \leq r(N)$. Let $A'_2 = a - N : a \in A_2$,

[1] This is not standard terminology.

[2] Actually he proved a stronger statement, namely, $\rho(N) \leq \frac{C}{\log \log N}$ for some C, i.e., $\rho(N) = O(\frac{1}{\log \log N})$.

[3] This is not standard notation.

so that $|A_2'| = |A_2|$, $A_2' \subseteq \{1, 2, \ldots, M\}$, and A_2' is good (since arithmetic progressions are preserved under translations)[4]. Thus $|A_2| \le r(M)$, so $r(N+M) = |A| = |A_1| + |A_2| \le r(N) + r(M)$, as desired. □

The following classical lemma is due to Pólya and Szegő [PS12] (part 1, problem 98), with a special case due to Fekete [Fek23].

Lemma 2. *If $(a_n)_{n\ge1}$ is a sequence of real numbers that satisfies*

$$a_{n+m} \le a_n + a_m \tag{1}$$

for all $n, m \ge 1$, and $\inf_{n\ge1} \frac{a_n}{n} \ne -\infty$, then $\frac{a_n}{n} \to \inf_{n\ge1} \frac{a_n}{n}$ as $n \to \infty$.

Proof. Define $\alpha = \inf_{n\ge1} \frac{a_n}{n}$, and let $\epsilon > 0$ be given. By definition of inf, there is some $m \ge 1$ such that $\frac{a_m}{m} < \alpha + \epsilon$. Let $n \ge 1$. Then by elementary number theory, there exist $q, r \in \mathbb{Z}, q, r \ge 0, r < m$ such that $n = mq + r$. Then, by repeated application of 1, we obtain $a_n = a_{qm+r} \le a_{qm} + a_r \le qa_m + a_r$, so

$$\frac{a_n}{n} = \frac{a_{qm_r}}{qm+r} \le \frac{qa_m + a_r}{qm+r} = \frac{qa_m}{qm+r} + \frac{a_r}{qm+r} \le \frac{qa_m}{qm} + \frac{a_r}{n} = \frac{a_m}{m} + \frac{a_r}{n}.$$

Thus

$$\alpha \le \frac{a_n}{n} \le \frac{a_m}{m} + \frac{a_r}{n} < \alpha + \epsilon + \frac{a_r}{n}.$$

Taking lim sup, since $|a_r| \le \max_{0\le k<m} |a_k|$ and the RHS does not depend on n, we obtain $\alpha \le \limsup_{n\to\infty} \frac{a_n}{n} < \alpha + \epsilon$, so

$$\limsup_{n\to\infty} \frac{a_n}{n} = \alpha. \tag{2}$$

By the properties of lim inf and lim sup, we have $\limsup_{n\to\infty} \frac{a_n}{n} \ge \liminf_{n\to\infty} \frac{a_n}{n} \ge \alpha$. Combining this with 2, we obtain $\lim_{n\to\infty} \frac{a_n}{n} = \alpha$, as claimed. □

Observe that since $\frac{r(N)}{N} \ge 0$ for all N, by combining Lemmas 1 and 2, we have $\rho(N) \to \inf_{n\ge1} \rho(N)$. Define $\rho = \inf_{n\ge1} \rho(N)$. For $N \ge 1$, let A_N be the lexicographically smallest good subset of $\{1, 2, \ldots, N\}$ of size $|A_N| = r(N)$. We define $S_N(x) = \sum_{a\in A_N} e(ax) = \sum_{k=1}^{N} 1_{A_N}(k)e(kx)$, $T_N(x) = \sum_{k=1}^{N} e(kx)$, $I_N = \int_0^1 S_N(x)^2 S_N(-2x)dx$ and $I_N^0 = \int_0^1 S_N(x)^2 T_N(-2x)dx$. We record several facts in the following lemma:

Lemma 3. *We have the following:*

$$r(N) = \int_0^1 |S_N(x)|^2 dx \tag{3}$$

$$r(N) = I_N \tag{4}$$

$$I_N^0 \ge \frac{r(N)^2}{4} \tag{5}$$

Proof. The proof is direct calculation.

$$\int_0^1 |S_N(x)|^2 dx = \int_0^1 S_N(x)\overline{S_N(x)}dx = \int_0^1 \sum_{a\in A_N} e(ax) \sum_{b\in A_N} \overline{e(bx)}dx \tag{6}$$

$$= \sum_{a,b\in A_N} \int_0^1 e((a-b)x)dx = \sum_{a,b\in A_N} \delta(a-b) = |A_N| = r(N), \tag{7}$$

[4]We record the following remark: Arithmetic progressions are preserved under translations and scalings, so $r(N)$ is equal to the size of the largest good subset of any arithmetic progression of length N. We will use this fact freely.

so 3 is proved.

$$I_N = \int_0^1 S_N(x)^2 S_N(-2x)dx = \int_0^1 \sum_{a\in A_N} e(ax) \sum_{b\in A_N} e(bx) \sum_{c\in A_N} e(-2cx) \tag{8}$$
$$= \sum_{a,b,c\in A_N} \int_0^1 e((a+b-2c)x)dx \tag{9}$$
$$= \sum_{a,b,c\in A_N} \delta(a+b-2c) = |A_N| = r(N), \tag{10}$$

where in the penultimate step we used the fact that A_N is good, which establishes 4.

$$I_N^0 = \int_0^1 S_N(x)^2 T_N(-2x)dx = \int_0^1 \sum_{a\in A_n} e(ax) \sum_{b\in A_N} e(bx) \sum_{c=1}^{N} e(-2cx)dx \tag{11}$$
$$= \sum_{a,b\in A_N} \sum_{c=1}^{N} \int_0^1 e((a+b-2c)x)dx \tag{12}$$
$$= \sum_{a,b\in A_N} \sum_{c=1}^{N} \delta(a+b-2c) \tag{13}$$
$$= |\{(a,b,c) : a+b=2c, a,b\in A_N, c\in\{1,2,\dots,N\}\}| \tag{14}$$
$$= |\{(a,b) : a,b\in A_N, a\equiv b \pmod 2\}| \tag{15}$$

Let $A_N^{(0)}$ be the set of even elements of A_N and $A_N^{(1)}$ be the set of odd elements of A_N, and $r_0 = |A_N^{(0)}|$, $r_1 = A_N^{(1)}$, so that $r(N) = r_0 + r_1$. Then, by 15,

$$I_N^0 = r_0^2 + r_1^2 \geq \frac{1}{2}(r_0^2 + r_1^2) = \frac{1}{2}\frac{(r_0+r_1)^2 + (r_0-r_1)^2}{2} \geq \frac{1}{4}(r_0+r_1)^2 = \frac{r(N)^2}{4}, \tag{16}$$

which completes the proof. □

We will use the following inequality.

Lemma 4. *For $0 < x < \frac{\pi}{2}$ we have*

$$\frac{2x}{\pi} \leq \sin(x) \leq x. \tag{17}$$

Proof. Let $f(x) = \sin(x) - x$. Then $f(0) = 0$ and $f'(x) = \cos(x) - 1 \leq 0$, so $f(x) \leq 0$ for $x > 0$. This gives the inequality on the right. For the inequality on the left, let $g(x) = \frac{\sin(x)}{x}$. Then $g'(x) = \frac{x\cos(x)-\sin(x)}{x^2}$. Let $h(x) = x\cos(x) - \sin(x)$. Then $h(0) = 0$ and $h'(x) = -x\sin(x) \leq 0$ for $0 < x < \frac{\pi}{2}$, so $h(x) \leq 0$ for $0 < x < \frac{\pi}{2}$. Thus $g'(x) \leq 0$ for $0 < x < \frac{\pi}{2}$, so g is non-increasing on $(0, \frac{\pi}{2})$. Thus $g(x) \geq g(\frac{\pi}{2}) = \frac{2}{\pi}$, so $\sin(x) \geq \frac{2x}{\pi}$, as desired. □

The following lemma of Dirichlet will also be useful.

Lemma 5 (Dirichlet)**.** *Let $\alpha \in \mathbb{R}$. Then for all $A \in \mathbb{N}$, there exists a rational number $\frac{a}{q}$ such that $1 \leq q \leq A$ and $|\alpha - \frac{a}{q}| \leq \frac{1}{qA}$.*

Proof. Consider the A numbers

$$\{\alpha\}, \{2\alpha\}, \dots, \{A\alpha\},$$

and the $A+1$ intervals

$$[0, \frac{1}{A+1}), [\frac{1}{A+1}, \frac{2}{A+1}), \dots, [\frac{A}{A+1}, 1).$$

If some $\{q\alpha\}$ is in the first or last interval, then we are done, since there exists some $a \in \mathbb{Z}$ such that $|q\alpha - a| \leq \frac{1}{A+1} \leq \frac{1}{A}$, which gives the desired rational number after dividing both sides by q. Otherwise, we have A numbers in $A-1$ intervals, so by pigeonhole principle there exists an interval $[\frac{b}{A+1}, \frac{b+1}{A+1})$

which contains at least two of the numbers, say, $\{h_1\alpha\}$ and $\{h_2\alpha\}$, with $h_1 < h_2$. Take $q = h_2 - h_1$ and $a = [h_2\alpha] - [h_1\alpha]$. Then

$$|q\alpha - a| = |h_2\alpha - h_1\alpha - ([h_2\alpha] - [h_1\alpha])| = |\{h_2\alpha\} - \{h_1\alpha\}| \leq \frac{1}{A+1} \leq \frac{1}{A}. \tag{18}$$

Dividing by q, we obtain the desired result. □

3 The proof of the theorem

In this section we will prove that $\rho = 0$. In order to prove this, we introduce some functions, for which we will prove some bounds. These bounds will translate to an expression for ρ, which will prove the theorem.

We define

$$F_M(x) = \sum_{z=0}^{M-1} e(zx), \tag{19}$$

$$E_{N,M}(x) = \rho(M)T_N(x) - S_N(x), \tag{20}$$

$$\sigma_{M,q}(y) = \sum_{k=0}^{M-1} (\rho(M) - 1_{A_N}(y + kq)). \tag{21}$$

Note that F_M is 1-periodic and we have

$$|F_M(x)| = \left|\sum_{z=0}^{M-1} e(x)^z\right| = \left|\frac{1 - e(Mx)}{1 - e(x)}\right| = \left|\frac{1 - e^{2\pi iMx}}{1 - e^{2\pi ix}}\right| = \left|\frac{e^{\pi iMx}(e^{-\pi iMx} - e^{\pi iMx})}{e^{\pi ix}(e^{-\pi ix} - e^{\pi ix})}\right| \tag{22}$$

$$= \left|\frac{e^{-\pi iMx} - e^{\pi iMx}}{e^{-\pi ix} - e^{\pi ix}}\right| = \left|\frac{\frac{e^{-\pi iMx} - e^{\pi iMx}}{2i}}{\frac{e^{-\pi ix} - e^{\pi ix}}{2i}}\right| \tag{23}$$

$$= \left|\frac{\sin(\pi Mx)}{\sin(\pi x)}\right| \tag{24}$$

for $x \notin \mathbb{Z}$.

The following lemma is the key step of the proof.

Lemma 6. *If $q < \frac{N}{M}$, then for all $y \in \mathbb{N}$ with $1 \leq y \leq N - Mq$, we have the following:*

$$\sigma_{M,q}(y) \geq 0, \tag{25}$$

$$F_M(xq)E_{N,M}(x) = \sum_{y=1}^{N-Mq} \sigma_{M,q}(y)e(x(y + Mq - q)) + O^*(2m^2q) \tag{26}$$

Proof. We first prove 26. By definition, we have

$$F_M(xq)E_{N,M}(x) = \sum_{z=0}^{M-1} e(zqx) \sum_{k=1}^{N} (\rho(M) - 1_{A_N}(k))e(kx) \tag{27}$$

$$= \sum_{z=0}^{M-1} \sum_{k=1}^{N} (\rho(M) - 1_{A_N}(k))e((k + zq)x). \tag{28}$$

We make the change of variables $h = k + (z - (M-1)q)$, so that $k + zq = h + Mq - q$. Then

$$F_M(xq)E_{N,M}(x) = \sum_{h=1+q-Mq}^{N} e((h + Mq - q)x) \sum_{\substack{z=0 \\ 1 \leq h+(M-1-z)q \leq N}}^{M-1} (\rho(M) - 1_{A_N}(h + (M-1-z)q)). \tag{29}$$

Using the facts $1 \le h+(M-1-z)q \le N$ for all $1 \le h \le N-Mq$ and $0 \le z \le M-1$, $e((h+Mq-q)x) = O^*(1)$, each term in the inner sum is $O^*(1)$ and there are at most M terms, after splitting the sum we obtain

$$F_M(xq)E_{N,M}(x) = \sum_{h=1}^{N-Mq} e((h+Mq-q)x) \sum_{z=0}^{M-1} (\rho(M) - 1_{A_N}(h+(M-1-z)q)) \tag{30}$$

$$+ \sum_{h=1+q-Mq}^{0} O^*(M) + \sum_{h=N-Mq+1}^{N} O^*(M) \tag{31}$$

$$= \sum_{h=1}^{N-Mq} \sigma_{M,q}(h)e(x(h+Mq-q)) + (Mq-q+Mq)O^*(M) \tag{32}$$

$$= \sum_{y=1}^{N-Mq} \sigma_{M,q}(y)e(x(y+Mq-q)) + O^*(2M^2q), \tag{33}$$

as desired.

We now prove 25. Fix some y such that $1 \le y \le N - Mq$. Then we have

$$\sigma_{M,q}(y) = \sum_{k=0}^{M-1} (\rho(M) - 1_{A_N}(y+kq)) = r(M) - \sum_{k=0}^{M-1} 1_{A_N}(y+kq). \tag{34}$$

Let $R = \sum_{k=0}^{M-1} 1_{A_N}(y+kq)$. Then $R = |A_N \cap \{y, y+q, \dots, y+(M-1)q\}|$. The set on the RHS is a subset of A_N, so it is good. It is also a subset of an arithmetic progression of length M, so by a remark in a footnote of the previous section, $R \le r(M)$. Thus $\sigma_{M,q}(y) \ge 0$. □

The bound for $E_{N,M}(x)$ obtained in the following lemma will let us deduce almost immediately that $\rho = 0$.

Lemma 7. *For N, M satisfying $2M^2 < N$, for all $x \in \mathbb{R}$ we have*

$$|E_{N,M}(x)| \le \frac{\pi}{2}N(\rho(M) - \rho(N)) + 4\pi M^2. \tag{35}$$

Proof. By Lemma 5, there exists a rational number $\frac{a}{q}$ such that $1 \le q \le 2M$ and $\left|x - \frac{a}{q}\right| \le \frac{1}{2qM}$. Note that $Mq \le 2M^2 < N$, so $q < N/M$. Let $b = qx - a$. If $b = 0$, then $F_M(xq) = F_M(a) = F_M(0) = M$. Therefore we have

$$|E_{N,M}(x)| = \frac{M}{M}|E_{N,M}(x)| = \frac{1}{M}|F_M(xq)E_{N,M}(x)| \tag{36}$$

$$\le \frac{1}{M}\left(\sum_{y=1}^{N-Mq} \sigma_{M,q}(y) + 2M^2q\right) \tag{37}$$

$$\le \frac{1}{M}((F_M(0)E_{N,M}(0) - O^*(2M^2q)) + 2M^2q) \tag{38}$$

$$\le E_{N,M}(0) + 4Mq \tag{39}$$

$$\le N(\rho(M) - \rho(N)) + 8M^2, \tag{40}$$

where we used triangle inequality and 26 in 37, and 26 in reverse in 38.

If $b \ne 0$, then $0 < |b| < \frac{1}{2M}$ so $0 < |\pi M b| < \frac{\pi}{2}$. Since sin is an odd function, without loss of generality we may assume that $b > 0$. Then, by 24 and Lemma 4,

$$|F_M(xq)| = |F_M(xq-a)| = |F_M(b)| = \left|\frac{\sin(\pi Mx)}{\sin(\pi x)}\right| \tag{41}$$

$$\ge \frac{\sin(\pi M b)}{\pi b} \tag{42}$$

$$\ge M\frac{\sin(\pi M b)}{\pi M b} \ge \frac{2M}{\pi}. \tag{43}$$

Using 43, similarly to 40 we obtain

$$|E_{N,M}(x)| = \frac{\pi}{2M}\frac{2M}{\pi}|E_{N,M}(x)| \le \frac{\pi}{2M}|F_M(xq)E_{N,M}(x)| \tag{44}$$

$$\le \frac{\pi}{2M}\left(\sum_{y=1}^{N-Mq}\sigma_{M,q}(y) + 2M^2q\right) \tag{45}$$

$$\le \frac{\pi}{2M}((F_M(0)E_{N,M}(0) - O^*(2M^2q)) + 2M^2q) \tag{46}$$

$$\le \frac{\pi}{2}(E_{N,M}(0) + 4Mq) \tag{47}$$

$$\le \frac{\pi}{2}N(\rho(M) - \rho(N)) + 4\pi M^2. \tag{48}$$

Since $1 < \frac{\pi}{2}$ and $8 < 4\pi$, 40 is not greater than 48. Thus 35 holds for all x. □

We can now prove Roth's theorem.

Theorem 1 (Roth)**.** *We have*

$$\lim_{N\to\infty}\rho(N) = \rho = 0. \tag{49}$$

Proof. For $N, M \in \mathbb{N}$ with $2M^2 < N$, define $\Delta_{N,M} = |I_N - \rho(M)I_N^0|$. Then we have

$$\Delta_{N,M} = \left|\int_0^1 S_N(x)^2 S_N(-2x)dx - \rho(M)\int_0^1 S_N(x)^2 T_N(-2x)dx\right| \tag{50}$$

$$\le \int_0^1 |S_N(x)|^2|S_N(-2x) - \rho(M)T_N(-2x)|dx \tag{51}$$

$$\le \|E_{N,M}\|_\infty \int_0^1 |S_N(x)|^2dx \tag{52}$$

$$= \|E_{N,M}\|_\infty r(N) \tag{53}$$

$$\le r(N)\left(\frac{\pi}{2}N(\rho(M) - \rho(N)) + 4\pi M^2\right), \tag{54}$$

where we used 3 in 53 and 35 in 54. We also have

$$\Delta_{N,M} = |r(N) - \rho(M)I_N^0| \ge \rho(M)I_N^0 - r(N) \tag{55}$$

$$\ge \frac{\rho(M)r(N)^2}{4} - r(N) \tag{56}$$

$$= \frac{\rho(M)\rho(N)Nr(N)}{4} - r(N) \tag{57}$$

$$= r(N)\left(\frac{N\rho(M)\rho(N)}{4} - 1\right), \tag{58}$$

where we used 5 in 56, and the fact $r(N) = N\rho(N)$ in 57. Thus, combining 54 and 58, we obtain

$$\frac{N\rho(M)\rho(N)}{4} - 1 \le \frac{\pi}{2}N(\rho(M) - \rho(N)) + 4\pi M^2, \tag{59}$$

so

$$\rho(M)\rho(N) \le \frac{4}{N} + 2\pi(\rho(M) - \rho(N)) + \frac{16\pi M^2}{N}. \tag{60}$$

Keeping M fixed and taking limits as $N \to \infty$ (which is allowed since $2M^2 < N$ is not violated as $N \to \infty$), we obtain

$$\rho \cdot \rho(M) \le 2\pi(\rho(M) - \rho). \tag{61}$$

Taking limits as $M \to \infty$, we see that

$$\rho^2 \le 2\pi(\rho - \rho) = 0, \tag{62}$$

so $\rho = 0$. This completes the proof. □

References

[Fek23] Michael Fekete. "Über die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit ganzzahligen Koeffizienten". In: *Mathematische Zeitschrift* 17 (1923), pp. 228–249.

[KM23] Zander Kelley and Raghu Meka. *Strong Bounds for 3-Progressions.* 2023. arXiv: 2302.05537.

[OBr10] Kevin O'Bryant. *What is the shortest route to Roth's theorem?* MathOverflow. 2010. URL: https://mathoverflow.net/a/47117.

[PS12] George Pólya and Gabor Szegő. *Problems and Theorems in Analysis I. Series. Integral Calculus. Theory of Functions.* Springer Berlin, Heidelberg, 2012.

[Rot53] K. F. Roth. "On Certain Sets of Integers". In: *Journal of the London Mathematical Society* s1-28.1 (1953), pp. 104–109.

[Vau97] R. C. Vaughan. *The Hardy-Littlewood Method.* 2nd ed. Cambridge Tracts in Mathematics. Cambridge University Press, 1997.