CHAPTER 4

Roth's Theorem on Arithmetic Progressions

4.1. Introduction

A famous theorem of van der Waerden states that given any natural numbers ℓ and r, there exists $N_0(\ell, r)$ such that for every natural number $n > N_0(\ell, r)$, every partition of the set $\{1, 2, \ldots, n\}$ into r subsets will yield a subset which contains ℓ terms in arithmetic progression.

This result leads naturally to the following question. Suppose that $\mathcal{A}$ is a set of natural numbers. For every natural number $n \in \mathbb{N}$, let

$$A(n) = A(n, \mathcal{A}) = \sum_{\substack{a \in \mathcal{A} \\ a \leqslant n}} 1$$

and

$$D(n) = D(n, \mathcal{A}) = \frac{A(n)}{n};$$

in other words, $A(n)$ and $D(n)$ denote respectively the number and proportion of elements of the set $\{1, 2, \ldots, n\}$ that are also in $\mathcal{A}$. Define the upper asymptotic density of the set $\mathcal{A}$ by

$$\overline{d} = \overline{d}(\mathcal{A}) = \limsup_{n \to \infty} D(n).$$

Erdős and Turán conjectured that every set $\mathcal{A}$ of natural numbers with positive upper asymptotic density contains arbitrarily long arithmetic progrssions. This is equivalent to the statement that if there is a natural number ℓ such that the set $\mathcal{A}$ contains no arithmetic progression of ℓ terms, then $\overline{d}(\mathcal{A}) = 0$.

The Hardy–Littlewood method can be adapted to establish the case $\ell = 3$ of this conjecture, as demonstrated by Roth in the 1950's. The novelty of this approach is that the Hardy–Littlewood method is applied to study a sequence that is not explicitly given, such as k-powers of natural numbers or primes.

For every $n \in \mathbb{N}$, let

$$M(n) = \max\{|\mathcal{S}| : \mathcal{S} \subseteq \{1, 2, \ldots, n\},\ \mathcal{S} \text{ does not contain 3 terms in arithmetic progression}\},$$

where $|\mathcal{S}|$ denotes the number of elements of the set $\mathcal{S}$. In other words, $M(n)$ denotes the largest number of elements which can be taken from the set $\{1, 2, \ldots, n\}$ with no 3 of them in arithmetic progression. Also, for every $n \in \mathbb{N}$, let

$$\delta(n) = \frac{M(n)}{n}.$$

THEOREM 4.1. *Suppose that $n \in \mathbb{N}$ and $n \geqslant 3$. Then $\delta(n) \ll (\log \log n)^{-1}$.*

The Erdős–Turán conjecture is now known to be true for every positive integer ℓ, and is now universally known as Szemerédi's theorem. Szemerédi's proof is a tour de force in combinatorics, and does not use the Hardy–Littlewood technique.

Roth's technique involves working with a set $\mathcal{M} \subseteq \{1, 2, \ldots, n\}$ that satisfies $|\mathcal{M}| = M(n)$ and does not contain 3 terms in arithmetic progression. We keep this set $\mathcal{M}$ fixed throughout our discussion, and apply the Hardy–Littlewood technique on this set. More precisely, consider the generating function

$$f(\alpha) = \sum_{x \in \mathcal{M}} e(\alpha x). \tag{4.1}$$

Then

$$\begin{aligned} \int_0^1 f^2(\alpha) f(-2\alpha)\,\mathrm{d}\alpha &= \int_0^1 \sum_{x_1 \in \mathcal{M}} \sum_{x_2 \in \mathcal{M}} \sum_{x_3 \in \mathcal{M}} e(\alpha(x_1 + x_2 - 2x_3))\,\mathrm{d}\alpha \\ &= \sum_{x_1 \in \mathcal{M}} \sum_{x_2 \in \mathcal{M}} \sum_{x_3 \in \mathcal{M}} \int_0^1 e(\alpha(x_1 + x_2 - 2x_3))\,\mathrm{d}\alpha \\ &= \sum_{\substack{x_1 \in \mathcal{M}}} \sum_{\substack{x_2 \in \mathcal{M} \\ x_1 + x_2 = 2x_3}} \sum_{x_3 \in \mathcal{M}} 1 = M(n), \end{aligned} \tag{4.2}$$

since the only possible solutions of the equation

$$x_1 + x_2 = 2x_3, \quad x_1, x_2, x_3 \in \mathcal{M},$$

are the trivial solutions $x_1 = x_2 = x_3$.

The main idea of the proof of Theorem 4.1 is that if $M(n)$ were close to n, then the integral

$$\int_0^1 f^2(\alpha) f(-2\alpha)\,\mathrm{d}\alpha$$

would be close to $M^2(n)$, thus contradicting (4.2).

4.2. A Major Arc Type Argument

The first step of the argument is to approximate the generating function (4.1). This can be achieved with a relatively small error if we make use of the disorderly arithmetical structure of the set $\mathcal{M}$. Sums of the form

$$\sum_{\substack{x=1 \\ x \in \mathcal{A}}}^{n} e(\alpha x)$$

tend to have large modulus near rational points a/q if the elements of $\mathcal{A}$ are well distributed in residue classes modulo q.

More precisely, suppose that the natural number $m < n$. Write

$$v(\alpha) = \delta(m) \sum_{x=1}^{n} e(\alpha x) \quad \text{and} \quad E(\alpha) = v(\alpha) - f(\alpha). \tag{4.3}$$

If we let $\chi_{\mathcal{M}}$ denote the characteristic function of the set $\mathcal{M}$, then

$$f(\alpha) = \sum_{x} \chi_{\mathcal{M}}(x) e(\alpha x).$$

Hence

$$E(\alpha) = \sum_{x=1}^{n} c(x) e(\alpha x),$$

where

$$c(x) = \delta(m) - \chi_{\mathcal{M}}(x).$$

THEOREM 4.2. *Suppose that*

$$g(\alpha) = \sum_{z=0}^{m-1} e(\alpha z). \tag{4.4}$$

Suppose further that the natural number $q < n/m$. Then

$$g(\alpha q)E(\alpha) = \sum_{h=1}^{n-mq} \sigma(h)e(\alpha(h + mq - q)) + R(\alpha), \tag{4.5}$$

where, for every $h = 1, \dots, n - mq$,

$$\sigma(h) = \sum_{x=0}^{m-1} c(h + xq) \geqslant 0,$$

and where

$$|R(\alpha)| < 2m^2 q. \tag{4.6}$$

Proof. It is easy to see that

$$g(\alpha q)E(\alpha) = \sum_{z=0}^{m-1} \sum_{x=1}^{n} c(x)e(\alpha(x + qz)).$$

Note that $x + qz \in [1, n + mq - q]$. Writing $x + qz = h + mq - q$, we have

$$\begin{aligned} g(\alpha q)E(\alpha) &= \sum_{h=1+q-mq}^{n} e(\alpha(h + mq - q)) \sum_{\substack{z=0 \\ 1\leqslant h+mq-q-qz\leqslant n}}^{m-1} c(h + mq - q - qz) \\ &= \sum_{h=1}^{n-mq} e(\alpha(h + mq - q)) \sum_{\substack{z=0 \\ 1\leqslant h+q(m-1-z)\leqslant n}}^{m-1} c(h + q(m - 1 - z)) + R(\alpha), \end{aligned} \tag{4.7}$$

where

$$\begin{aligned} R(\alpha) = &\sum_{h=1+q-mq}^{0} e(\alpha(h + mq - q)) \sum_{\substack{z=0 \\ 1\leqslant h+q(m-1-z)\leqslant n}}^{m-1} c(h + q(m - 1 - z)) \\ &+ \sum_{h=n-mq+1}^{n} e(\alpha(h + mq - q)) \sum_{\substack{z=0 \\ 1\leqslant h+q(m-1-z)\leqslant n}}^{m-1} c(h + q(m - 1 - z)). \end{aligned} \tag{4.8}$$

Now the inner sums in (4.8) clearly do not exceed m in absolute value, so

$$|R(\alpha)| \leqslant m(mq - q + mq) < 2m^2 q. \tag{4.9}$$

On the other hand, if $1 \leqslant h \leqslant n - mq$, then for every integer z in the range $0 \leqslant z \leqslant m - 1$, the inequality $1 \leqslant h + q(m - 1 - z) \leqslant n$ is always satisfied. It follows from (4.7) that

$$g(\alpha q)E(\alpha) = \sum_{h=1}^{n-mq} \left(\sum_{z=0}^{m-1} c(h + q(m - 1 - z)) \right) e(\alpha(h + mq - q)) + R(\alpha), \tag{4.10}$$

where

$$\sum_{z=0}^{m-1} c(h + q(m - 1 - z)) = \sum_{x=0}^{m-1} c(h + xq) = \sigma(h). \tag{4.11}$$

The inequalities (4.5) and (4.6) now follow from (4.9)–(4.11). Note next that

$$\sigma(h) = \sum_{x=0}^{m-1} (\delta(m) - \chi_{\mathcal{M}}(h + xq)) = M(m) - \sum_{x=0}^{m-1} \chi_{\mathcal{M}}(h + xq). \tag{4.12}$$

The sum

$$r = \sum_{x=0}^{m-1} \chi_{\mathcal{M}}(h + xq)$$

is the number of elements of $\mathcal{M}$ in the arithmetic progression

$$h, h + q, \dots, h + (m - 1)q.$$

Let these elements be $h+x_1q,\dots,h+x_rq$. Now no three of these are in arithmetic progression. Hence no three of $x_1,\dots,x_r$ are in arithmetic progression, whence no three of $1+x_1,\dots,1+x_r$ are in arithmetic progression. Also $1+x_j \leqslant m$ for every $j=1,\dots,r$. It follows that we must have $r \leqslant M(m)$, whence $\sigma(h) \geqslant 0$, in view of (4.12). ◯

THEOREM 4.3. *Suppose that $2m^2 < n$. Then for every real number α, we have*

$$|E(\alpha)| < 2n(\delta(m)-\delta(n)) + 16m^2.$$

PROOF. By Dirichlet's theorem, there exist integers a and q satisfying $(a,q)=1$ and $1 \leqslant q \leqslant 2m$ such that

$$\left|\alpha - \frac{a}{q}\right| \leqslant \frac{1}{2mq}.$$

Then

$$g(\alpha q) = g(\alpha q - a) = g(\beta), \tag{4.13}$$

where

$$|\beta| = |\alpha q - a| \leqslant \frac{1}{2m}.$$

It follows from (4.4) and (4.13) that

$$|g(\alpha q)| = |g(\beta)| = \left|\frac{\sin \pi m\beta}{\sin \pi\beta}\right| \geqslant \frac{2m}{\pi}.$$

Note next that $q \leqslant 2m < n/m$. In view of Theorem 4.2, we have

$$\begin{aligned} \frac{m}{2}|E(\alpha)| \leqslant \frac{2m}{\pi}|E(\alpha)| \leqslant |g(\alpha q)E(\alpha)| &< \sum_{h=1}^{n-mq} \sigma(h) + 2m^2 q \\ &= g(0)E(0) - R(0) + 2m^2q < mE(0) + 4m^2q \leqslant mE(0) + 8m^3. \end{aligned} \tag{4.14}$$

On the other hand,

$$E(0) = \sum_{x=1}^{n}(\delta(m) - \chi_{\mathcal{M}}(x)) = n\delta(m) - M(n) = n(\delta(m)-\delta(n)). \tag{4.15}$$

The result follows on combining (4.14) and (4.15). ◯

4.3. Completion of the Proof

Write

$$I = \int_0^1 f^2(\alpha)v(-2\alpha)\,\mathrm{d}\alpha. \tag{4.16}$$

In view of (4.1) and (4.3), we have

$$\begin{aligned} I &= \int_0^1 \sum_{x_1\in\mathcal{M}}\sum_{x_2\in\mathcal{M}}\sum_{y=1}^{n}\delta(m)e(\alpha(x_1+x_2-2y))\,\mathrm{d}\alpha \\ &= \sum_{x_1\in\mathcal{M}}\sum_{x_2\in\mathcal{M}}\sum_{y=1}^{n}\delta(m)\int_0^1 e(\alpha(x_1+x_2-2y))\,\mathrm{d}\alpha \\ &= \sum_{x_1\in\mathcal{M}}\sum_{x_2\in\mathcal{M}}\sum_{\substack{y=1\\x_1+x_2=2y}}^{n}\delta(m) = \sum_{\substack{x_1\in\mathcal{M}\ x_2\in\mathcal{M}\\ x_1+x_2\ \text{even}}}\delta(m). \end{aligned}$$

Let M_1 and M_2, where $M_1+M_2=M(n)$, denote respectively the number of odd and even elements of $\mathcal{M}$. Then

$$I = \delta(m)(M_1^2+M_2^2) \geqslant \frac{1}{2}\delta(m)(M_1+M_2)^2 = \frac{1}{2}\delta(m)M^2(n). \tag{4.17}$$

On the other hand, it follows from (4.2), (4.3) and (4.16) that

$$|M(n) - I| = \left| \int_0^1 f^2(\alpha)(f(-2\alpha) - v(-2\alpha))\, \mathrm{d}\alpha \right| \leqslant \left(\max_\alpha |E(\alpha)| \right) \int_0^1 |f(\alpha)|^2 \, \mathrm{d}\alpha.$$

Clearly

$$\int_0^1 |f(\alpha)|^2 \, \mathrm{d}\alpha = \int_0^1 f(\alpha) f(-\alpha) \mathrm{d}\alpha = M(n).$$

It follows from Theorem 4.3 that if $2m^2 < n$, then

$$|M(n) - I| \leqslant (2n(\delta(m) - \delta(n)) + 16m^2) M(n). \tag{4.18}$$

Combining (4.17) and (4.18), we have

$$\frac{1}{2} n M(n) \delta(m) \delta(n) = \frac{1}{2} \delta(m) M^2(n) \leqslant I \leqslant M(n) + (2n(\delta(m) - \delta(n)) + 16m^2) M(n),$$

so that

$$\delta(m)\delta(n) \leqslant 2n^{-1} + 4(\delta(m) - \delta(n)) + 32m^2 n^{-1} \leqslant 4(\delta(m) - \delta(n)) + 34m^2 n^{-1}, \tag{4.19}$$

so long as $2m^2 < n$.

THEOREM 4.4. *The limit*

$$\tau = \lim_{n \to \infty} \delta(n) \tag{4.20}$$

exists. Furthermore, $\delta(n_2) \leqslant 2\delta(n_1)$ *for all natural numbers* $n_1 \leqslant n_2$.

PROOF. It is trivial that $M(m + n) \leqslant M(m) + M(n)$. Hence for $n_2 \geqslant n_1$,

$$M(n_2) = M\left(n_1 \left[\frac{n_2}{n_1} \right] + \left(n_2 - n_1 \left[\frac{n_2}{n_1} \right] \right) \right) \leqslant \left[\frac{n_2}{n_1} \right] M(n_1) + M\left(n_2 - n_1 \left[\frac{n_2}{n_1} \right] \right). \tag{4.21}$$

Clearly

$$M(n_2) \leqslant \frac{n_2}{n_1} M(n_1) + n_1,$$

so that

$$\delta(n_2) \leqslant \delta(n_1) + \frac{n_1}{n_2}.$$

Hence

$$\limsup_{n_2 \to \infty} \delta(n_2) \leqslant \delta(n_1) \quad \text{and} \quad \limsup_{n_2 \to \infty} \delta(n_2) \leqslant \liminf_{n_1 \to \infty} \delta(n_1),$$

so the limit (4.20) exists. Also, it follows from (4.21) that

$$M(n_2) \leqslant \frac{n_2}{n_1} M(n_1) + M(n_1) \leqslant 2 \frac{n_2}{n_1} M(n_1).$$

The second assertion follows immediately. ○

REMARK. Letting $n \to \infty$, the inequality (4.19) becomes

$$\delta(m)\tau \leqslant 4(\delta(m) - \tau).$$

Letting $m \to \infty$, we conclude that $\tau^2 \leqslant 0$, so that $\tau = 0$. This is a weaker form of Theorem 4.1.

To complete the proof of Theorem 4.1, we write

$$\lambda(x) = \delta(2^{3^x}).$$

In view of Theorem 4.4, it suffices to prove that $\lambda(x) \ll x^{-1}$. By (4.19), we have

$$\lambda(y)\lambda(y+1) \leqslant 4(\lambda(y) - \lambda(y+1)) + 34 \cdot 2^{-3^y},$$

so that

$$1 \leqslant \frac{4(\lambda(y) - \lambda(y+1))}{\lambda(y)\lambda(y+1)} + \frac{34 \cdot 2^{-3^y}}{\lambda(y)\lambda(y+1)}.$$

Summing this over $y = x, x+1, \ldots, 2x-1$, we have

$$\begin{aligned} x &\leqslant \sum_{y=x}^{2x-1} \frac{4(\lambda(y) - \lambda(y+1))}{\lambda(y)\lambda(y+1)} + \sum_{y=x}^{2x-1} \frac{34 \cdot 2^{-3^y}}{\lambda(y)\lambda(y+1)} \\ &= 4 \sum_{y=x}^{2x-1} \left(\frac{1}{\lambda(y+1)} - \frac{1}{\lambda(y)} \right) + \sum_{y=x}^{2x-1} \frac{34 \cdot 2^{-3^y}}{\lambda(y)\lambda(y+1)} \\ &\leqslant \frac{4}{\lambda(2x)} + \frac{200x2^{-3^x}}{\lambda^2(2x)}, \end{aligned}$$

in view of Theorem 4.4. When $\lambda(2x) > 1/x$, then

$$\frac{200x2^{-3^x}}{\lambda^2(2x)} < \frac{x}{2}$$

for all sufficiently large x, so that $\lambda(2x) < 8/x$.