Acta Math. Hung.
56 (1—2) (1990), 155—158.

INTEGER SETS CONTAINING NO ARITHMETIC PROGRESSIONS

E. SZEMERÉDI (Budapest), corresponding member of the Academy

Introduction

K. F. Roth [1] proved 1953 using analytic methods that if a strictly increasing sequence of natural numbers $a_1 < a_2 < \ldots < a_k \leqq n$ contains no three term arithmetic progression then

$$k < \frac{c_1 n}{\log\log n}. \tag{1}$$

Very recently Heat—Brown [2] could improve considerably (1) by showing

$$k < \frac{c_2 n}{(\log n)^{c_3}} \quad (c_3 > 0). \tag{2}$$

The aim of the present work is to show that Roth's analytic method combined with some combinatorial ideas is are useful in the study of such type problems. Applying the method to the present problem the resulting inequality will be (2) whilst in [3] it was shown that if $a_1 < a_2 < \ldots < a_k \leqq n$ is a sequence $\mathcal{A}$ of natural numbers such that $\mathcal{A} - \mathcal{A}$ does not contain any positive square then

$$k < \frac{c_4 n}{(\log n)^{\log\log\log\log n/12}}.$$

It is possible that the present approach leads to new results in other problems of additive number theory too.

NOTATIONS. Let

$$\mathcal{N}_{(n)} = \{1, 2, \ldots, n\},$$

$$\mathcal{N}_{i,j,q,s} = \{iq+j, (i+1)q+j, (i+s-1)q+j\},$$

$$\mathcal{A}_{i,j,q,s} = \mathcal{N}_{i,j,q,s} \cap \mathcal{A}.$$

Let $|\mathcal{A}|$ be the number of elements in $\mathcal{A}$,

$$f_{\mathcal{A}}(\alpha) = \sum_{a \in \mathcal{A}} e(a\alpha),\ e(\alpha) = e^{2\pi i\alpha}, \quad \gamma = \frac{|\mathcal{A}|}{n}.$$

Let ε be a sufficiently small positive number and $n > n_0(\varepsilon)$. We suppose $\gamma > \frac{c_2}{\log^{c_3} n}$ otherwise the theorem is trivially true for $\mathcal{A}$.

Let us assume the assertion is proved for every $m \leqq \sqrt{n}$. Choosing c_2 sufficiently small it is clearly true for $n < c_4$. It is easy to see that either

a) $\mathcal{A}$ has a subset $\mathcal{A}'\subset\left[H, H+\frac{n}{\log n^3}\right]$ with

$$|\mathcal{A}'|>(1+\varepsilon)\gamma\frac{n}{(\log n)^3},\quad 1\leqq H\leqq n.$$

or

b) the total number of solutions of the equations $a_i+a_j=2a_k+n_0$ and $a_i+a_j=$ $=2a_k-n_0$ is at most

$$\frac{1}{2}(1+4\varepsilon)\gamma^3 n_0^2 \tag{3}$$

for some number $n_0\in\left[n, n+\frac{n}{\log^2 n}\right]$. The case a) can be settled easily (cf. the end of the proof). Now we are dealing with the case b). In the following we work with n_0 instead of n.

We write $\gamma_0=\gamma\cdot\frac{n}{n_0}$. Now

$$\frac{1}{n_0}\sum_{t=0}^{n_0-1} f_{\mathcal{A}}\left(\frac{t}{n_0}\right)f_{\mathcal{A}}\left(\frac{t}{n_0}\right)f_{\mathcal{A}}\left(\frac{-2t}{n_0}\right)$$

is the number of solutions of the equation $a_i+a_j\equiv 2a_k \pmod{n_0}$. In view of (3) this is less than $\left(\frac{1}{2}+3\varepsilon\right)\gamma_0^3 n_0^2$, since there is no 3-term arithmetic progression in $\mathcal{A}$.

Because the main term (corresponding to $t=0$) is $\gamma_0^3\cdot n_0^3$ it follows that

$$\frac{1}{n_0}\left|\sum_{t=1}^{n_0-1} f_{\mathcal{A}}\left(\frac{t}{n_0}\right)f_{\mathcal{A}}\left(\frac{t}{n_0}\right)f_{\mathcal{A}}\left(\frac{-2t}{n_0}\right)\right|>\left(\frac{1}{2}-3\varepsilon\right)\frac{\gamma_0^3 n_0^2}{2}.$$

Let us assume that for $t\neq 0$, $\left|f\left(\frac{t}{n_0}\right)\right|<\frac{|\mathcal{A}|}{2^{i_0}}$ with a fixed i_0, sufficiently large. There must be an $i=i_1$ with $2^{i_0}<2^{i_1}<(\log n)^{1/3}$ such that there exist $t_1, t_2, \dots, t_q$ $\left(q=q(i_1)=\left[\frac{2^{3i_1}}{i_1^2}\right]+1\right)$, $t_\mu\not\equiv t_\nu \pmod{n_0}$ with $\left|f\left(\frac{t_\mu}{n_0}\right)\right|>\frac{|\mathcal{A}|}{2^{i_1}}$. Otherwise we would have

$$\frac{1}{n_0}\sum_{\substack{i\geqq i_0\\ 2^i\leqq\log^{1/3}n}}\ \sum_{|\mathcal{A}|\cdot 2^{-i}\leqq\max\left(\left|f_{\mathcal{A}}\left(\frac{t}{n_0}\right)\right|,\left|f_{\mathcal{A}}\left(-\frac{2t}{n_0}\right)\right|\right)\leqq|\mathcal{A}|\cdot 2^{-i+1}}\left|f_{\mathcal{A}}\left(\frac{t}{n_0}\right)\right|^2\left|f_{\mathcal{A}}\left(-\frac{2t}{n_0}\right)\right|<\frac{|\mathcal{A}|^3}{10n_0}$$

while it follows from Parseval's identity that

$$\frac{1}{n_0}\sum_{\max\left(\left|f_{\mathcal{A}}\left(\frac{t}{n_0}\right)\right|,\left|f_{\mathcal{A}}\left(-\frac{2t}{n_0}\right)\right|\right)<|\mathcal{A}|\log^{-1/3}n}\left|f_{\mathcal{A}}\left(\frac{t}{n_0}\right)\right|^2\left|f_{\mathcal{A}}\left(-\frac{2t}{n_0}\right)\right|\ll$$

$$\ll|\mathcal{A}|\cdot\frac{|\mathcal{A}|}{\log^{1/3}n}=o\left(\frac{|\mathcal{A}|^3}{n_0}\right).$$

We shall show the existence of a set

$$\mathscr{B} = \{b, 2b, \dots, |\mathscr{B}|b\} = \{b_k\}_{k=1}^{|\mathscr{B}|}$$

such that

$$1 \leqq b \leqq n_0^{q/(q+1)}, \quad |\mathscr{B}| = \left[\frac{n^{1/(q+1)}}{(\log n)^2}\right],$$

$$jbt_v \equiv l_{i,j} \pmod{n_0}, \quad |l_{i,j}| < \frac{2n}{(\log n)^2}$$

for all $1 \leqq j \leqq |\mathscr{B}|, 1 \leqq v \leqq q = q(i_1)$. Dividing the set $\mathscr{N}_{(n_0)}$ into $n_0^{1/(q+1)}$ equal intervals $I_1, I_2, \dots, I_{n_0^{1/(q+1)}}$, there must exist b' and b'' $(1 \leqq b' < b'' \leqq n_0^{q/(q+1)})$ such that $b't_v \pmod{n_0}$ lies in the same interval as $b''t_v$ for all v (i.e. $|(b'-b'')t_v - k_v n_0| \leqq n_0^{q/(q+1)}$ with integer k_v). The choise $b = |b''-b'|$ satisfies our requirements.

Now the number of solutions of $a_i - a_j \equiv b_k - b_l \pmod{n_0}$ is

$$\frac{1}{n_0}\sum_{t=0}^{n_0-1} f_{\mathscr{A}}\left(\frac{t}{n_0}\right)\overline{f_{\mathscr{A}}\left(\frac{t}{n_0}\right)} f_{\mathscr{B}}\left(\frac{t}{n_0}\right)\overline{f_{\mathscr{B}}\left(\frac{t}{n_0}\right)}$$

which is at least $(i = i_1, n > n_0(\varepsilon))$

$$(1-\varepsilon)\frac{1}{n_0}\frac{|\mathscr{A}|^2}{2^{2i}}|\mathscr{B}|^2 \cdot \frac{2^{3i}}{i^2} = (1-\varepsilon)\gamma_0^2 n_0 |\mathscr{B}|^2 \cdot \frac{2^i}{i^2}. \tag{4}$$

On the other hand the number of solutions of $a_i - a_j \equiv b_k - b_l \pmod{n_0}$ (with the notation $B' = \frac{|\mathscr{B}|}{T}$, T a large constant) is

$$\left(1+\delta(T)\right)\sum_{j=1}^{b}\sum_{i=0}^{n/B'} |\mathscr{A}_{B'i,j,b,B'}| \sum_{h=-T}^{T} |\mathscr{A}_{B'(i+h),j,b,B'}| \left(1 - \frac{|h|}{T}\right)|\mathscr{B}| \tag{5}$$

where $\delta(T) \to 0$ as $T \to \infty$.

There exists a set $\mathscr{A}_{B'v,j,b,B'}$ with

$$|\mathscr{A}_{B'v,j,b,B'}| \geqq (1-2\varepsilon)\frac{2^i}{i^2}B'\gamma_0 \tag{6}$$

since otherwise the sum in (4) would be with a fixed $T = T_0(\varepsilon)$

$$\leqq (1+\varepsilon)|\mathscr{A}|T \cdot \frac{2^i}{i^2}(1-2\varepsilon)B'\gamma_0|\mathscr{B}| < (1-\varepsilon)\frac{2^i}{i^2}\gamma_0^2 n_0 |\mathscr{B}|^2$$

in contradiction to (5).

Similarly if we have a t^* with

$$\left|f_{\mathscr{A}}\left(\frac{t^*}{n_0}\right)\right| > \frac{|\mathscr{A}|}{2^{i_0}}$$

then the same argument (with $q=1$) shows the existence of a

$\frac{\sqrt{n_0}}{\log^2 n} > B' > \sqrt{n_0}/\log^3 n$ with

$$|\mathcal{A}_{B'\nu,j,b,B'}| > (1-2\varepsilon)\left(1+\frac{1}{2^{2i_0}}\right)B'\gamma_0. \tag{6'}$$

In both cases (6) and (6'), using the set $\mathcal{A}_{B'\nu,j,b,B'}$ we obtain a new set $\mathcal{A}' \subset \subset \{1, \ldots, B'\}$ with $B' \in \left[\frac{n^{1/(q+1)}}{\log^3 n}, n^{1/(q+1)}\right]$ such that $\mathcal{A}'$ contains no 3 terms arithmetic progressions and $|\mathcal{A}'| > c(q)B'\gamma_0$. Here either

i) $q = 1,\ c(q) > 1 + \frac{1}{2^{2i_0+1}}$

or

ii) $q > \frac{2^{3i_0}}{i_0^2},\ c(q) > \frac{q^{1/3}}{\log^{4/3} q}$.

It is easy to see that we have in both cases

$$|\mathcal{A}'| > c_2 \frac{c(q)}{\log^{c_3} n} B' > \frac{c_2 B'}{\log^{c_3} B'}$$

if c_3 was chosen sufficiently small. This contradicts our induction hypothesis and so proves the theorem,

REMARK. It is easy to see that if $\left|f_{\mathcal{A}}\left(\frac{t}{n}\right)\right| < \frac{|\mathcal{A}|}{2^{i_0}}$ for all $t = 1, 2, \ldots, n-1$ with a sufficiently large i_0, (i.e. we have case b) preceding (3)) then c_0 can be chosen near 1/3. ($c_0 = 1/3 - \varepsilon$ is admissible if $i_0 > c_0(\varepsilon)$). A more careful computation concerning case a) shows that $c_0 > 1/4$ can be chosen in the formulation of the Theorem.

References

[1] K. F. Roth, On certain sets of integers, *J. London Math. Soc.*, **28** (1953), 104—109.

[2] D. R. Heath—Brown, Integer sets containing no arithmetic progressions (to appear).

[3] J. Pintz, W. L. Steiger and E. Szemerédi, On sets of natural numbers whose difference set contains no squares, to appear in *J. London Math. Soc.*

(Received November 16, 1987)

MATHEMATICAL INSTITUTE
OF THE HUNGARIAN ACADEMY OF SCIENCES
H—1053 BUDAPEST, REÁLTANODA U. 13—15.