Combinatorics, Probability and Computing (2006) 15, 597–603.
doi:10.1017/S0963548306007516 Printed in the United Kingdom

Arithmetic Progressions in Sets with Small Sumsets

JÓZSEF SOLYMOSI†

Department of Mathematics, University of British Columbia,
1984 Mathematics Road, Vancouver, BC, Canada V6T 1Z2
(e-mail: solymosi@math.ubc.ca)

Received 11 January 2004; revised 24 August 2004

We present an elementary proof that if A is a finite set of numbers, and the sumset $A +_G A$ is small, $|A +_G A| \leqslant c|A|$, along a dense graph G, then A contains k-term arithmetic progressions.

1. Introduction

A well-known theorem of Szemerédi [15] states that every dense subset of integers contains long arithmetic progressions. A different, but somehow related result of Freiman [5] says that if the sumset of a finite set of numbers A is small, *i.e.*, $|A + A| \leqslant C|A|$, then A is the subset of a (not very large) generalized arithmetic progression. Balog and Szemerédi proved in [1] that a similar structural statement holds under weaker assumptions. (For correct statements and details, see [8].) As a corollary of their result, Freiman's theorem, and Szemerédi's theorem about k-term arithmetic progressions, Balog and Szemerédi proved Theorem 1.1 below. The goal of this paper is to present a simple, purely combinatorial proof of this assertion.

Let A be a set of numbers and G be a graph such that the vertex set of G is A. The *sumset of A along G* is

$$A +_G A = \{a + b : a, b \in A \text{ and } (a, b) \in E(G)\}.$$

Theorem 1.1. *For every $c, K, k > 0$ there is a threshold $n_0 = n_0(c, K, k)$ such that if $|A| = n \geqslant n_0$, $|A +_G A| \leqslant K|A|$, and $|E(G)| \geqslant cn^2$, then A contains a k-term arithmetic progression.*

† Research partially supported by NSERC and OTKA grants.

2. Lines and hyperplanes

There are arrangements of n lines on the Euclidean plane such that the maximum number of points incident with at least three lines is $\frac{n^2}{6}$. Not much is known about the structure of arrangements where the number of such points is close to the maximum, say cn^2, where c is a positive constant. Nevertheless, the following is true.

Lemma 2.1. *For every $c > 0$ there is a threshold $n_0 = n_0(c)$ and a positive $\delta = \delta(c)$ such that, for any set of $n \geqslant n_0$ lines L and any set of $m \geqslant cn^2$ points P, if every point is incident to three lines, then there are at least δn^3 triangles in the arrangement.* (*A triangle is a set of three distinct points from P such that any two are incident to a line from L.*)

Proof. This lemma is implied by the following theorem of Ruzsa and Szemerédi [13].

Theorem 2.2. ([13]) *Let G be a graph on n vertices. If G is the union of cn^2 edge-disjoint triangles, then G contains at least δn^3 triangles, where δ depends on c only.*

To prove Lemma 2.1, let us construct a graph where L is the vertex set, and two vertices are adjacent if and only if the corresponding lines cross at a point of P. This graph is the union of cn^2 disjoint triangles, and every point of P defines a unique triangle, so we can apply Theorem 2.2. □

The result above suffices to prove Theorem 1.1 for 3-term arithmetic progressions. But for larger values of k, we need a generalization of Lemma 2.1.

Lemma 2.3. *For every $c > 0$ and $d \geqslant 2$, there is a threshold $n_0 = n_0(c, d)$ and a positive $\delta = \delta(c, d)$ such that, for any set of $n \geqslant n_0$ hyperplanes L and any set of $m \geqslant cn^d$ points P, if every point is incident to $d + 1$ hyperplanes, then there are at least δn^{d+1} simplices in the arrangement.* (*A simplex is a set of $d + 1$ distinct points from P such that any d are incident to a hyperplane from L.*)

Lemma 2.3 follows from the Frankl–Rödl conjecture [4], the generalization of Theorem 2.2. The $d = 3$ case was proved in [4] and the conjecture has been proved recently by Gowers [6] and independently by Nagle, Rödl, Schacht and Skokan [7, 10]. For details on how Lemma 2.3 follows from the Frankl–Rödl conjecture, see [14].

3. The $k = 3$ case

Let A be a set of numbers and G be a graph such that the vertex set of G is A. We define the *difference-set of A along G* as

$$A -_G A = \{a - b : a, b \in A \text{ and } (a, b) \in E(G)\}.$$

Lemma 3.1. *For every $\epsilon, c, K > 0$ there is a number $D = D(\epsilon, c, K)$ such that, if $|A +_G A| \leqslant K|A|$ and $|E(G)| \geqslant c|A|^2$, then there is a graph $G' \subset G$ such that $|E(G')| \geqslant (1 - \epsilon)|E(G)|$ and $|A -_{G'} A| \leqslant D|A|$.*

Proof. Let us consider the arrangement of points given by a subset of the Cartesian product $A \times A$ and the lines $y = a$, $x = a$ for every $a \in A$, and $x + y = t$ for every $t \in A +_G A$. The pointset P is defined by $(a, b) \in P$ if and only if $(a, b) \in E(G)$. By Lemma 2.1, the number of triangles in this arrangement is δn^3. The triangles here are right isosceles triangles. We say that a point in P is *popular* if the point is the right-angle vertex of at least αn triangles. Selecting $\alpha = \frac{\delta(\epsilon c)}{\epsilon c}$, where $\delta(\cdot)$ is the function from Lemma 2.1, all but at most $\epsilon c n^2$ points of P are popular.

A $t \in A - A$ is *popular* if $|\{(a, b) : a - b = t; a, b \in A\}| \geqslant \alpha n$. The number of popular ts is at most Dn, where D depends on α only. $A \times A$ is a Cartesian product; therefore every triangle can be extended to a square adding one extra point from $A \times A$. Every popular point p is the right-angle vertex of at least αn triangles. Therefore p is incident to a line $x - y = t$, where t is popular, because this line contains at least αn 'fourth' vertices of squares with p. □

Proof of Theorem 1.1, case $k = 3$. Let us apply Lemma 2.1 to the pointset P' defined by $(a, b) \in P'$ if and only if $(a, b) \in E(G')$ and the lines are $y = a$ for every $a \in A$, $x - y = t$ for every $t \in A -_{G'} A$, and $x + y = s$ for every $s \in A +_G A$. By Theorem 2.2, if $|A|$ is large enough, then there are triangles in the arrangement. The vertices of such triangles are vertices from $P' \subset A \times A$. The vertical lines through the vertices form a 3-term arithmetic progression and therefore A contains δn^2 3-term arithmetic progressions, where $\delta > 0$ depends on c only. □

4. The general, $k > 3$, case

Following the steps of the proof for $k = 3$, we prove the general case by induction on k. We prove the following theorem, which was conjectured by Erdős, and proved by Balog and Szemerédi in [1]. Theorem 4.1, together with the $k = 3$ case, gives a proof of Theorem 1.1.

Theorem 4.1. *For every $c > 0$ and $k > 3$ there is an n_0 such that, if A contains at least $c|A|^2$ 3-term arithmetic progressions and $|A| \geqslant n_0$, then A contains a k-term arithmetic progression.*

Instead of triangles, we must consider simplices. Set $k = d$. In the d-dimensional space we show that $A \times \cdots \times A$, the d-fold Cartesian product of A, contains a simplex in which the vertices' first coordinates form a $(d + 1)$-term arithmetic progression.

The simplices we are looking for are homothetic[1] images of the simplex S_d whose vertices are listed below:

$$
\begin{gathered}
(0,0,0,0,\ldots,0,0)\\
(1,1,0,0,\ldots,0,0)\\
(2,0,1,0,\ldots,0,0)\\
(3,0,0,1,\ldots,0,0)\\
\vdots\\
(d-1,0,\ldots,1,0)\\
(d,0,0,0,\ldots,0,0).
\end{gathered}
$$

[1] Here we say that two simplices are homothetic if the corresponding facets are parallel.

An important property of S_d is that its facets can be pushed into a 'shorter' grid. The facets of S_d are parallel to hyperplanes, defined by the origin $(0,0,0,0,\ldots,0,0)$, and some $(d-1)$-tuples of the grid

$$\{0,1,2,\ldots,d-1\} \times \{-1,0,1\} \times \{0,1\}^{d-2}.$$

For example, if $d = 3$, then the facets are

$$\begin{gathered}\{(0,0,0),(1,1,0),(2,0,1)\}\\ \{(0,0,0),(1,1,0),(3,0,0)\}\\ \{(0,0,0),(2,0,1),(3,0,0)\}\\ \{(1,1,0),(2,0,1),(3,0,0)\},\end{gathered}$$

and the corresponding parallel planes in

$$\{0,1,2\} \times \{-1,0,1\} \times \{0,1\}$$

are the planes incident to the triples

$$\begin{gathered}\{(0,0,0),(1,1,0),(2,0,1)\}\\ \{(0,0,0),(1,1,0),(2,0,0)\}\\ \{(0,0,0),(2,0,1),(2,0,0)\}\\ \{(0,0,0),(1,-1,1),(2,-1,0)\}.\end{gathered}$$

In general, if a facet of S_d contains the origin and the 'last point' $(d,0,0,0,\ldots,0,0)$, then if we replace the later one by $(d-1,0,0,0,\ldots,0,0)$, the new d-tuples define the same hyperplane. The remaining facet f, given by

$$\begin{gathered}(1,1,0,0,\ldots,0,0)\\ (2,0,1,0,\ldots,0,0)\\ (3,0,0,1,\ldots,0,0)\\ \vdots\\ (d-1,0,\ldots,1,0)\\ (d,0,0,0,\ldots,0,0),\end{gathered}$$

is parallel to the hyperplane through the vertices of $f-(1,1,0,0,\ldots,0,0)$,

$$\begin{gathered}(0,0,0,0,\ldots,0,0)\\ (1,-1,1,0,\ldots,0,0)\\ (2,-1,0,1,\ldots,0,0)\\ \vdots\\ (d-2,-1,\ldots,1,0)\\ (d-1,-1,0,0,\ldots,0,0).\end{gathered}$$

In a homothetic copy of the grid

$$T_d = \{0,1,2,\ldots,d-1\} \times \{-1,0,1\} \times \{0,1\}^{d-2},$$

the image of the origin is called the *holder* of the grid.

As the induction hypothesis, let us suppose that Theorem 4.1 is true for a $k \geqslant 3$ in a stronger form, provided that the number of k-term arithmetic progressions in A is at least $c|A|^2$.

Then the number of distinct homothetic copies of T_d in

$$\mathbb{A}_d = \underbrace{A \times \ldots \times A}_{d}$$

is at least $c'|A|^{d+1}$ (c' depends on c only). Let us say that a point $p \in \mathbb{A}_d$ is *popular* if p is the holder of at least $\alpha|A|$ grids. If p is popular, then, for any facet of S_d, f, the point p is the element of at least $\alpha|A|$ d-tuples, similar and parallel to f. If α is small enough, then at least $\gamma|A|^d$ points of $\mathbb{A}_d$ are popular, where γ depends on c and α only.

A hyperplane H is *β-rich* if it is incident to many points, $|H \cap \mathbb{A}_d| \geqslant \beta|A|^{d-1}$. For every facet of S_d, f, let us denote the set of β-rich hyperplanes which are parallel to f by $\mathscr{H}_f$.

Lemma 4.2. *For some choice of β, at least half of the popular points are incident to $d+1$ β-rich hyperplanes, parallel to the facets of S_d.*

Suppose to the contrary that, for a facet f, more than $\frac{\gamma}{2d}|A|^d$ popular points are not incident to hyperplanes of $\mathscr{H}_f$. Then, more than

$$\alpha|A|\frac{\gamma}{2d}|A|^d = \frac{\gamma\alpha}{2d}|A|^{d+1} \tag{4.1}$$

d-tuples, similar and parallel to f, are not covered by $\mathscr{H}_f$. Let us denote the hyperplanes incident to the 'uncovered' d-tuples by $L_1, L_2, \ldots, L_m$, and the number of points on the hyperplanes by $\mathscr{L}_1, \mathscr{L}_2, \ldots, \mathscr{L}_m$. A simple result of Elekes and Erdős [2, 3] implies that hyperplanes with few points cannot cover many d-tuples.

Theorem 4.3. ([3]) *The number of homothetic copies of f in L_i is at most $c_d\mathscr{L}_i^{1+1/(d-1)}$, where c_d depends on d only.*

The inequalities

$$\sum_{i=1}^{m} \mathscr{L}_i \leqslant |A|^d, \text{ and } \mathscr{L}_i \leqslant \beta|A|^{d-1}.$$

lead us to the proof of Lemma 4.2.

The number of d-tuples covered by L_is is at most

$$c_d \sum_{i=1}^{m} \mathscr{L}_i^{1+1/(d-1)} \leqslant c_d \frac{|A|^d}{\beta|A|^{d-1}} (\beta|A|^{d-1})^{1+1/(d-1)} = c_d\beta^{1/(d-1)}|A|^{d+1}.$$

If we compare this bound to (4.1), and choose β such that $\frac{\gamma\alpha}{2d} = c_d\beta^{1/(d-1)}$, then at least half of the popular points are covered by $d+1$ β-rich hyperplanes parallel to the facets of S_d.

Finally we can apply Lemma 2.3 with the pointset P of 'well-covered' popular points of $\mathbb{A}_d$ and with the sets of hyperplanes $L = \bigcup_{f \subset S_d} \mathscr{H}_f$. The number of points is at least

$\frac{\gamma\alpha}{2}|A|^d$. For a given f, $|\mathcal{H}_f| \leqslant \frac{|A|^d}{\beta|A|^{d-1}} = |A|/\beta$. The number of hyperplanes in L is at most $(d+1)|A|/\beta$. By Lemma 2.3, we have at least $\delta'|A|^{d+1}$ homothetic copies of S_d in $\mathbb{A}_d$. Let us project them onto x_1, the first coordinate axis. Every image is a $(k+1)$-term arithmetic progression, and the multiplicity of one image is at most $|A|^{d-1}$. Therefore there are at least $\delta'|A|^2$ $(k+1)$-term arithmetic progressions in A.

5. $G_n = K_n$

When the full sumset $A+A$ is small then it is easier to prove that A contains long arithmetic progressions. We can use the following Plünnecke-type inequality [8, 9, 12].

Theorem 5.1. *Let A and B be finite subsets of an abelian group such that $|A| = n$ and $|A+B| \leqslant \delta n$. Let $k \geqslant 1$ and $l \geqslant 1$. Then*

$$|kB - lB| \leqslant \delta^{k+l} n.$$

It follows from the inequality that, for any dimension d and d-dimensional integer vector $\vec{v} = (x_1, \ldots, x_d)$, $x_i \in \mathbb{Z}$, there is a $c > 0$ depending on d, δ and $\vec{v}$ such that the following holds: *If $|A+A| \leqslant \delta|A|$, then $\mathbb{A}_d$ can be covered by $c|A|$ hyperplanes with the same normal vector $\vec{v}$.* Using this, we can define our hyperplane-point arrangement, with the hyperplanes parallel to the facets of S_d containing at least one point of $\mathbb{A}_d$, and the pointset of the arrangement is $\mathbb{A}_d$. Then we do not have to deal with rich planes and popular points, and we can apply Lemma 2.3 directly.

References

[1] Balog, A. and Szemerédi, E. (1994) A statistical theorem of set addition. *Combinatorica* **14** 263–268.

[2] Elekes, G. (2002) Sums versus products in number theory, algebra and Erdős geometry. In *Paul Erdős and his Mathematics II*, Budapest, Vol. 11 of *Bolyai Society Mathematical Studies*, p. 277.

[3] Elekes, G. and Erdős, P. (1994) Similar configurations and pseudo grids. In *Intuitive Geometry*, Vol. 63 of *Coll. Math. Soc. János Bolyai*, North-Holland, Amsterdam, pp. 85–104.

[4] Frankl, P. and Rödl, V. (2002) Extremal problems on set systems. *Random Struct. Alg.* **20** 131–164.

[5] Freiman, G. A. (1973) *Foundations of Structural Theory of Set Addition*, Vol. 37 of *Translation of Mathematical Monographs*, AMS, Providence, RI, USA.

[6] Gowers, W. T. Hypergraph regularity and the multidimensional Szemerédi theorem. Manuscript.

[7] Nagle, B., Rödl, V. and Schacht, M. The counting lemma for regular k-uniform hypergraphs. *Random Struct. Alg.* **28** 113–179.

[8] Nathanson, M. B. (1996) *Additive Number Theory: Inverse Problems and the Geometry of Sumsets*, Vol. 165 of *Graduate Texts in Mathematics*, Springer.

[9] Plünnecke, H. (1969) *Eigenschaften und Abschätzungen von Wirkungsfunctionen*, Vol. 22, Berichte der Gesellschaft für Mathematik und Datenverarbeitung, Bonn.

[10] Rödl, V. and Skokan, J. (2004) Regularity lemma for k-uniform hypergraphs. *Random Struct. Alg.* **25** 1–42.

[11] Roth, K. F. (1953) On certain sets of integers. *J. London Math. Soc.* **28** 245–252.

[12] Ruzsa, I. Z. (1989) An application of graph theory to additive number theory. *Scientica Ser. A* **3** 97–109.

[13] Ruzsa, I. Z. and Szemerédi, E. (1978) Triple systems with no six points carrying three triangles. In Vol. 18 of *Colloquia Mathematica Societatis János Bolyai* (Proc. 5th Hungarian Colloq., Keszthely, Hungary, 1976), North-Holland, pp. 939–945.
[14] Solymosi, J. (2004) A note on a question of Erdős and Graham. *Combin. Probab. Comput.* **13** 263–267.
[15] Szemerédi, E. (1975) On sets of integers containing no k elements in arithmetic progression. *Acta Arithmetica* **27** 199–245.