INTEGERS: ELECTRONIC JOURNAL OF COMBINATORIAL NUMBER THEORY 7(2) (2007), #A10

ARITHMETIC PROGRESSIONS IN SPARSE SUMSETS

Ernie Croot[1]
School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332

Imre Ruzsa [2]
Alfréd Rényi Institute of Mathematics, Budapest, pf. 127, H-1364 Hungary

Tomasz Schoen [3]
Department of Discrete Mathematics, Adam Mickiewicz University, ul. Umultowska 87, 61-614 Poznań, Poland

Received: 11/29/05, Accepted: 12/15/05

Abstract

In this paper we show that sumsets $A+B$ of finite sets A and B of integers, must contain long arithmetic progressions. The methods we use are completely elementary, in contrast to other works, which often rely on harmonic analysis.

–Dedicated to Ron Graham on the occasion of his 70^{th} *birthday*

1. Introduction

Given a set C of an additive group G, we let $L(C)$ denote the length of the longest arithmetic progression in C, where given the arithmetic progression $a, a+d, a+2d, ..., a+(k-1)d$ of distinct elements in G, we define the length of this progression to be k.

One of the main focuses in combinatorial (and additive) number theory is that of understanding the structure of the sumset $2A := A + A = \{a + b : a, b \in A\}$, given certain information about the set A. For example, one such problem is to determine $L(2A)$, given

[1]Supported by an NSF grant
[2]Supported by Hungarian National Foundation for Scientific Research (OTKA) Grants T38396, T42750, and T43623
[3]Partially supported by grant 1 P03A 029 30

that $A \subseteq [N] := \{1, 2, ..., N\}$ and $|A| > \delta N$, for some $0 < \delta \leq 1$. The first major progress on this problem was due to J. Bourgain [1], who proved the beautiful result:

Theorem 1 *If $A, B \subseteq [N]$ and $|A| = \gamma N$ and $|B| = \delta N$, then for N large enough,*

$$L(A + B) \ > \ \exp[c(\gamma\delta \log N)^{1/3} - \log\log N],$$

for some constant c.

Then, I. Ruzsa [7] gave a construction, which is the following theorem:

Theorem 2 *For every $\epsilon > 0$ and every sufficiently large prime p, there exists a symmetric set A of residues modulo p (i.e. $A = -A$) with $|A| \geq p(1/2 - \epsilon)$, such that*

$$L(2A) \ < \ \exp((\log p)^{2/3+\epsilon}).$$

A simple consequence of this theorem is that for N sufficiently large, there exists a set $A \subset [N]$ with $|A| > (1/2 - \epsilon)N$, such that

$$L(2A) \ < \ \exp((\log N)^{2/3+\epsilon}),$$

which shows that the $1/3$ in Bourgain's result cannot be improved to any number beyond $2/3$.

In a recent paper, B. Green [4] proved the following beautiful result, which improves upon Bourgain's result above, and is currently the best that is known on this problem:

Theorem 3 *Suppose A, B are subsets of $\mathbb{Z}/N\mathbb{Z}$ having cardinalities γN and δN, respectively. Then there is an absolute constant $c > 0$ such that*

$$L(A + B) \ > \ \exp(c((\gamma\delta \log N)^{1/2} - \log\log N)).$$

There are also several other papers which treat the question of long arithmetic progressions in sumsets $A + A + \cdots + A$, such as [3], [5], [6], [9], [10], [11], and [12].

In this paper we give a proof of a result, which shows that sumsets $2A$ have long arithmetic progressions when $A \subseteq [N]$ has only $N^{1-\theta}$ elements (the length of the longest progression will depend on θ). This result is stronger than those given in the above theorems of Bourgain and Green when $|A|, |B| \ll N(\log N)^{-1/2}$; however, when $|A|, |B| > N(\log N)^{-1/2+\epsilon}$, their results give a much stronger conclusion.

First, we need some notation: We define $\mathrm{odd}(n)$ to be the smallest odd integer that is $\geq n$; so, $n \leq \mathrm{odd}(n) < n + 2$. Our first theorem is as follows.

Theorem 4 *Suppose that $A \subset \mathbb{Z}$, and that*

$$|A - A| \;=\; C|A|, \text{ and } |A - 2A| \;=\; K|A|. \tag{1}$$

Then,

$$L(A - A) \;\geq\; \text{odd}\left(2\frac{\log|A|}{\log K} + 1\right). \tag{2}$$

$$L(2A) \;\geq\; \text{odd}\left(2\frac{\log(C^{-1}|A|)}{\log(CK)} \;+\; 1\right). \tag{3}$$

$$L(2A) \;\geq\; \text{odd}\left(\frac{\log(C^{-1}|A|)}{2\log C} + 1\right). \tag{4}$$

A corollary of this theorem is as follows:

Corollary 1 *For every odd $k \geq 1$ and N sufficiently large, if*

$$A \;\subseteq\; [N], \text{ and } |A| \geq (3N)^{1-1/(k-1)},$$

then $L(2A) \geq k$.

Also, if

$$A, B \;\subseteq\; [N], \text{ and } |A||B| \;\geq\; 6N^{2-2/(k-1)},$$

then $L(A + B) \geq k$.

To compare this result with those of Bourgain and Green, we note that when $|A|, |B| \gg N$, then Green's result gives that $A+B$ contains a progression of length $\exp(c(\log N)^{1/2})$, for some constant c, whereas the authors' result above gives only $\Omega(\log N)$. So, in this range, both Green's and Bourgain's results are much stronger than Theorem 4 and its corollary; however, when $|A|, |B| \ll N/\sqrt{\log N}$, then Green's result does not give a non-trivial bound on the length of the longest arithmetic progression in $A + B$, whereas the result above gives that $A + B$ contains a progression of length $\Omega((\log N)/\tau \log\log N)$ when

$$|A|, |B| \;\gg\; \frac{N}{\log^{\tau} N},$$

for any $\tau > 0$. Another point is that in Theorem 4 and its corollary, the arithmetic progressions produced contain 0, whereas the arithmetic progressions in Green's result do not.

We also have a construction of sets A such that $2A$ has no long arithmetic progressions. This construction is the following theorem:

Theorem 5 *For every $\epsilon > 0$, there exists $0 < \theta_0 \leq 1$ so that if $0 < \theta < \theta_0 \leq 1$, then there exist infinitely many integers N and sets $A \subseteq [N]$ with $|A| \geq N^{1-\theta}$, such that*

$$L(2A) \ < \ \exp(c\theta^{-2/3-\epsilon}),$$

where $c > 0$ is some absolute constant.

The rest of the paper is organized as follows: In the next section we will present some open problems on arithmetic progressions in sumsets; and, in the last section, we will give proofs of all the theorems listed above.

2. Open Questions

From Theorem 5 and Corollary 1 we deduce that for every $\epsilon > 0$ and $0 < \theta < 1$ sufficiently small,

$$\frac{2}{\theta} + O(1) \ < \ \min_{\substack{A \subseteq [N] \\ |A| \geq N^{1-\theta}}} L(2A) \ < \ \exp(c\theta^{-2/3-\epsilon}). \qquad (5)$$

This brings us to the following, difficult problem:

Problem 1. What is the true size of $\min_{A \subseteq [N], |A| = N^{1-\theta}} L(2A)$?

Another way to look at problems concerning arithmetic progressions is to fix the length k of the progression, and to determine the parameter θ guaranteeing a k-term arithmetic progression. This problem (which is just a restatement of Problem 1) is as follows:

Problem 2. Fix $k \geq 1$. Given N, determine the largest $\theta \in (0, 1)$ such that if $A \subseteq [N]$ satisfies $|A| \geq N^{1-\theta}$, then $L(2A) \geq k$.

One can interpret (5) as saying that this largest $\theta = \theta(N)$ satisfies

$$\frac{2}{k} \ \ll \ \theta \ \ll_\epsilon \ \frac{1}{(\log k)^{3/2-\epsilon}}.$$

for all N sufficiently large.

In the case $k = 3$ we have from Corollary 1 that if $|A| > N^{1-\theta}$, $A \subseteq [N]$, and $\theta > 1/2 + O(1/\log N)$, then $2A$ contains a three-term arithmetic progression. On the other hand, if A is a B_4 set, which is a set containing no non-trivial solutions to

$$x_1 + x_2 + x_3 + x_4 \ = \ x_5 + x_6 + x_7 + x_8, \ x_1, ..., x_8 \in A,$$

then $2A$ contains no three-term progressions, since in particular it contains no solutions to

$$(x_1 + x_2) + (x_3 + x_4) \ = \ 2(x_5 + x_6).$$

Now, it is known from [2] that B_4 sets with more than $N^{1/4}$ elements exist for N sufficiently large. Thus, we have in the special case $k = 3$, in partial answer to Problem 2, the largest θ for which $|A| \geq N^{1-\theta}$ implies $L(2A) \geq 3$ satisfies

$$\frac{1}{2} + O\left(\frac{1}{\log N}\right) \ < \ \theta \ \leq \ \frac{3}{4},$$

for N sufficiently large.

3. Proofs of Theorems and Corollaries

Proof of Theorem 4.

Define m to be the largest integer satisfying

$$m \ < \ \frac{\log |A|}{\log K} + 1, \tag{6}$$

and assume that (1) holds. Since $A - A$ is symmetric and contains 0, we have that (2) holds if

$$d, 2d, ..., md \ \in \ A - A, \tag{7}$$

since this would imply that

$$-md, -(m-1)d, ..., 0, d, ..., md \ \in \ A - A,$$

which has length $2m + 1$.

Now, (7) holds if and only if $d = a_1 - b_1 \in A - A$ and

$$a_{j+1} - b_{j+1} \ = \ a_j - b_j + a_1 - b_1, \text{ for } j = 1, ..., m-1, \tag{8}$$

(Here, all $a_j - b_j \in A - A$) if and only if $d = a_1 - b_1$ and

$$a_{j+1} - a_j - a_1 \ = \ b_{j+1} - b_j - b_1, \text{ for } j = 1, ..., m-1.$$

If we had two sequences $a_1, ..., a_m$ such that the derived sequences $a_{j+1} - a_j - a_1$ coincide, we have a solution to (8). Now, let V denote the set of all vectors of length $m - 1$ given by

$$(a_2 - 2a_1, \ a_3 - a_2 - a_1, \ a_4 - a_3 - a_1, \ ..., \ a_m - a_{m-1} - a_1).$$

We note that since each coordinate here lies in $A - 2A$, we have from (1) that

$$|V| \ \leq \ K^{m-1}|A|^{m-1}.$$

Thus, since there are $|A|^m$ choices for $a_1, ..., a_m$, we have that (8) has a solution if

$$|A|^m \ > \ |V| \ = \ K^{m-1}|A|^{m-1};$$

in other words,

$$|A| \ > \ K^{m-1}.$$

This inequality holds because m satisfies (6), and so we have proved (2).

To prove (3), we observe from the Cauchy-Schwarz inequality that

$$\sum_{a,b \in A} |(a-A) \cap (A-b)| \ = \ \sum_{n \in A-A} w(n)^2 \ \geq \ |A|^4 |A-A|^{-1}.$$

where $w(n)$ is the number of ways of writing $n = a - b$, $a, b \in A$. Thus, from (1) we have that for some $a, b \in A$ if we let $B = A \cap (a + b - A)$, then

$$|B| \ \geq \ C^{-1}|A|,$$

and

$$B - B \ \subseteq \ 2A - a - b.$$

It follows that

$$|B - 2B| \ \leq \ |A - 2A| \ = \ K|A| \ \leq \ CK|B|,$$

and so

$$\begin{aligned} L(2A) \ \geq \ L(B-B) \ &\geq \ \text{odd}\left(2\frac{\log|B|}{\log CK} + 1\right) \\ &\geq \ \text{odd}\left(2\frac{\log(C^{-1}|A|)}{\log CK} + 1\right). \end{aligned}$$

Thus, we have proved (3).

Finally, to prove (4) we apply the following result due to Ruzsa [8, Lemma 3.3].

Lemma 1 *Suppose that A is a subset of an additive group G, and that*

$$|A - A| \ \leq \ H|A|.$$

Then,

$$|A \pm A \pm A \cdots \pm A| \ \leq \ H^t|A|,$$

where t is the number of terms here.

From this lemma, we deduce that if

$$|A - A| \ \leq \ C|A|,$$

then

$$|A - 2A| \ \le \ C^3|A|,$$

and so, $K \le C^3$ and it follows from (3) that

$$L(2A) \ \ge \ \text{odd}\left(\frac{\log(C^{-1}|A|)}{2\log C} + 1\right).$$

Proof of the Corollary 1.

Since $A - A$ is a subset of $\{-N+1, ..., N-1\}$, which has size $2N-1$, we have that

$$C \ = \ \frac{|A-A|}{|A|} \ < \ \frac{2}{3}(3N)^{1/(k-1)}. \tag{9}$$

Also, since

$$|2A - A| \ \le |\{-N+2, ..., 2N-1\}| \ < \ 3N,$$

we deduce

$$K \ < \ (3N)^{1/(k-1)}. \tag{10}$$

From (3) we deduce that

$$\begin{aligned} L(2A) \ &\ge \ \text{odd}\left(2\frac{\log(C^{-1}|A|)}{\log(CK)} + 1\right) \\ &\ge \ \text{odd}\left(2\frac{\log(3(3N)^{1-2/(k-1)}/2)}{\log(2(3N)^{2/(k-1)}/3)} + 1 + \epsilon\right) \\ &= \ \text{odd}(k - 2 + \epsilon_1) \\ &\ge \ k, \end{aligned}$$

where $\epsilon_1 > 0$ is some constant, and comes from the fact that (9) and (10) are strict inequalities.

For every pair $(a, b) \in A \times B$ there exists a unique $t \in [2, 2N]$ such that $a = t - b$. Thus,

$$\sum_{2 \le t \le 2N} |A \cap (t - B)| \ = \ |A||B|,$$

and it follows that there exists an integer t such that if we set $D = A \cap (t - B)$, then

$$|D| \ \ge \ \frac{|A||B|}{2N-1} \ > \ 3N^{1-2/(k-1)}.$$

Since

$$D - D + t \ \subseteq \ A + B,$$

and since

$$|D - 2D| \ \le \ |[1-2N, N-1]| \ = \ 3N - 1 \ < \ N^{2/(k-1)}|D|,$$

we have from (2) (applied with the set D) that

$$\begin{aligned} L(A+B) \ \geq \ L(D-D) \quad & \geq \quad \mathrm{odd}\left(\frac{2\log|D|}{\log(N^{2/(k-1)})} + 1 + \epsilon_2(k,N) \right) \\ & \geq \quad \mathrm{odd}(k-2+\epsilon_2) \\ & \geq \quad k, \end{aligned} \tag{11}$$

where $\epsilon_2 > 0$ is some constant depending on N and k.

Proof of Theorem 5.

From Theorem 2 we have that for every $\epsilon > 0$, there exists $0 < \theta < 1$ so that if we let

$$K \ = \ \left\lfloor 10^{\theta^{-1}} \right\rfloor + 1, \tag{12}$$

then there exists a set $S \subseteq \{0, ..., K-1\}$ satisfying $|S| \geq (K-1)(1/2-\epsilon) > K/5$, and

$$L(S+S) \ < \ \exp((\log K)^{2/3+\epsilon}).$$

Given such a set S, define A to be the set of all integers of the form

$$a_0 + a_1(2K) + a_2(2K)^2 + \cdots + a_{t-1}(2K)^{t-1}, \text{ where } a_i \in S,$$

where $t \geq 1$ is arbitrary. Let $N = (2K)^t$, and note that $A, 2A \subset \{0, ..., N\}$.

Now, we have that, regardless of what value we choose for $t \geq 1$,

$$|A| \ \geq \ \left(\frac{K}{5}\right)^t \ > \ (2K)^{t(1-\theta)} \ = \ N^{1-\theta}.$$

The last inequality here follows from (12).

We also have that

$$\begin{aligned} L(2A) \ = \ L(S+S) \quad & < \quad \exp((\log K)^{2/3+\epsilon}) \\ & < \quad \exp(c\theta^{-2/3-\epsilon}), \end{aligned}$$

for some constant $c > 0$.

References

[1] J. Bourgain, *On Arithmetic Progressions in Sums of Sets of Integers*, A tribute to Paul Erdős, 105-109, Cambridge University Press, Cambridge, 1990.

[2] J. Cilleruelo and J. Himénez-Urroz, $B_h[g]$ *sequences*, Mathematika **47** (2000), 109-115.

[3] G. A. Freiman, H. Halberstam, and I. Ruzsa, *Integer Sums sets Containing Long Arithmetic Progressions*, J. London Math. Soc. (2) **46** (1992), 193-201.

[4] B. Green, *Arithmetic Progressions in Sumsets*, Geom. Funct. Anal. **12** (2002), 584-597.

[5] V. F. Lev, *Optimal Representations by Sumsets and Subset Sums*, J. Number Theory **62** (1997), 127-143.

[6] ———-, *Blocks and Progressions in Subset Sums Sets*, Acta Arith. **106** (2003), 123-142.

[7] I. Ruzsa, *Arithmetic Progressions in Sumsets*, Acta. Arith. **60** (1991), no. 2, 191-202.

[8] ———, *Arithmetic Progressions and the Number of Sums*, Periodica Math. Hung. **33** (1992), 105-111.

[9] A. Sárközy, *Finite Addition Theorems. I*, J. Number Theory **32** (1989), 114-130.

[10] ————, *Finite Addition Theorems. II*, J. Number Theory **48** (1994), 197-218.

[11] ————, *Finite Addition Theorems. III*, Groupe de Travail en Théorie Analytique et Élémentaire des Nombres, 1989-1990, 105-122, Publ. Math. Orsay, 92-01, Univ. Paris XI, Orsay, 1992.

[12] E. Szemerédi and V. Vu, *Long Arithmetic Progressions in Sumsets and x-Sum-Free-Sets*, J. London Math. Soc. **3** (2005), no. 2, 273-296.