ON A CONJECTURE OF ALON

LINH TRAN, VAN VU, AND PHILIP MATCHETT WOOD

Abstract. Let $f(n, m)$ be the cardinality of largest subset of $\{1, 2, \dots, n\}$ which does not contain a subset whose elements sum to m. In this note, we show that

$$f(n, m) = (1 + o(1))\frac{n}{\mathrm{snd}(m)}$$

for all $n(\log n)^{1+\epsilon} \le m \le \frac{n^2}{9\log^2 n}$, where $\mathrm{snd}(m)$ is the smallest integer that does not divide m. This proves a conjecture of Alon posed in [1].

1. Introduction

For n a large positive integer and m an integer between n and n^2, we define $f(n, m)$ to be the maximum cardinality of a set $A \subset \{1, 2, \dots, n\}$ such that no subset $B \subset A$ satisfies $\sum_{b \in B} b = m$. In 1986, Erdős and Graham [4] observed that $f(n, m) \ge \left(\frac{1}{2} + o(1)\right)\frac{n}{\log n}$. (Here, and throughout this paper, log denotes the natural logarithm unless otherwise specified, so $\log x := \log_e x$. The asymptotic notation is used under the assumption that $n \to \infty$.)

For s a positive integer not dividing m, it is clear that $f(n, m) \ge \left\lfloor \frac{n}{s} \right\rfloor$, since any sum of elements of the set $\left\{s, 2s, 3s, 4s, \dots, \left\lfloor \frac{n}{s} \right\rfloor s\right\}$ cannot divide (and hence cannot equal) m. Letting $\mathrm{snd}(m)$ denote the smallest positive integer that does not divide m, we thus have

$$\left\lfloor \frac{n}{\mathrm{snd}(m)} \right\rfloor \le f(n, m). \tag{1}$$

By the prime number theorem, we know that $\mathrm{snd}(m) \le (2 + o(1))\log n$, and so (1) matches the lower bound observed by Erdős and Graham [4]. In 1987, Alon [1] made the following conjecture, which essentially states that the lower bound is asymptotically sharp.

Conjecture 1.1. *If $n^{1.1} \le m \le n^{1.9}$, then*

$$f(n, m) = (1 + o(1))\frac{n}{\mathrm{snd}(m)}.$$

There have been several partial results concerning this conjecture. In [1], Alon (using extremal graph theory, a theorem due to Moser and Scherk [8], and Roth's Theorem [11]) proved

Theorem 1.2. [1] *For every $\epsilon > 0$ there exists a constant $c = c(\epsilon) \ge 1$ such that for every n and*

$$n^{1+\epsilon} \le m \le \frac{n^2}{\log^2 n},$$

we have

Date: April 26, 2009.

The authors are partially supported by NSF Grant 0635606.

$$\left\lfloor \frac{n}{\operatorname{snd}(m)} \right\rfloor \le f(n,m) \le \frac{cn}{\operatorname{snd}(m)}.$$

Later, Lipkin [7] (using analytic methods along the lines of those in [5]) showed

Theorem 1.3. [7] *There exist positive constants c and C such that the following holds for all positive integers n and m. If*

$$cn\log^6 n < m < \frac{n^{3/2}}{\log^3 n},$$

then

$$f(n,m) \le \frac{n}{\operatorname{snd}(m)} + C\frac{n\log(\operatorname{snd}(m))}{\operatorname{snd}(m)\log^2 n} = (1+o(1))\frac{n}{\operatorname{snd}(m)}.$$

In another paper, Alon and Freiman [2] (again using analytic methods) determined the precise value of $f(n,m)$ for large m,

Theorem 1.4. [2] *For every $\epsilon > 0$ there is a constant $n_0 = n_0(\epsilon)$ such that the following holds. If $n \ge n_0$ and*

$$3n^{5/3+\epsilon} < m < \frac{n^2}{20\log^2 n},$$

then

$$f(n,m) = \left\lfloor \frac{n}{\operatorname{snd}(m)} \right\rfloor + \operatorname{snd}(m) - 2.$$

In this note, we prove Conjecture 1.1 in full using a theorem of Sárközy (see Theorem 2.1) and elementary arguments.

Theorem 1.5. *For any constants $c > 0$ and $\epsilon > 0$, there is a constant $n_0 = n_0(c,\epsilon)$ such that the following holds. If $n \ge n_0$ and*

$$cn(\log n)^{1+\epsilon} \le m \le \frac{n^2}{9\log^2 n},$$

then

$$f(n,m) = (1+o(1))\frac{n}{\operatorname{snd}(m)}.$$

Our methods can be used to prove the following *inverse* result, which characterizes the structure of relatively large sets A where no subsum sums up to m. Similar results have been obtained for finite fields (see [16, 9, 10] or [18] for a survey), but the arguments here are quite different. This result essentially says that the example giving the lower bound in (1) is the only way for a reasonably large subset of $\{1, 2, \ldots, n\}$ to avoid containing a subset that sums up to m.

Theorem 1.6. *Let c, δ, ϵ_1, and ϵ_2 be positive constants such that $0 < \epsilon_1 < \epsilon_2$, and let m and n be integers satisfying*

$$cn(\log n)^{1+\epsilon_2} \le m \le \frac{\delta^2 n^2}{8(\log n)^{2+2\epsilon_1}},$$

where we assume that n is sufficiently large. If

$$\frac{\delta n}{(\log n)^{1+\epsilon_1}} \le |A|$$

and if no subset $B \subset A$ satisfies $\sum_{b\in B} b = m$, then A contains $(1-o(1))\,|A|$ elements that are congruent to $0 \mod d$*, where d is an integer that does not divide m.*

2. Long arithmetic progressions in iterated sumsets

Given a set A of integers, we define

$$\begin{aligned} \ell A &:= \{a_1 + a_2 + \cdots + a_\ell : a_i \in A\}, \\ \ell^* A &:= \{a_1 + a_2 + \cdots + a_\ell : \text{ the } a_i \text{ are distinct elements of } A\}, \text{ and} \\ S_A &:= \left\{ m : \text{ there exists } B \subset A \text{ satisfying } \sum_{b\in B} b = m \right\}. \end{aligned}$$

Notice that $\ell^* A \subset S_A$.

The key fact that lets us prove Theorem 1.5 is that iterated sumsets ℓA and $\ell^* A$ exhibit more and more arithmetic structure as ℓ increases, and they even exhibit substantial structure for relatively small values of ℓ. The first results on arithmetic progressions in ℓA were produced by Freiman, Halberstam, and Ruzsa [6], by Bourgain [3], and by Sárközy [12]. Later results in this direction also applied to $\ell^* A$, for example those of Sárközy [13, 14] and recently those of Szemerédi and Vu [16, 15, 17].

The main tool we will use is the following result due to Sárközy [14].

Theorem 2.1. [14] *Let $n \in \mathbb{N}$ be such that $n > 2500$, let $A' \subset \{1, 2, \ldots, n\}$, and say*

$$|A'| > 100\sqrt{n \log n}.$$

Then, for every $L \in \mathbb{N}$ such that

$$n \le L \le \frac{10^{-4}\,|A'|^2}{\log(13n/\,|A'|)},$$

there exists d, ℓ, and L_0 such that

$$\begin{aligned} 1 &\le d \le \frac{4828n}{|A'|}, \\ 1 &\le \ell \le \frac{8496L}{|A'|}, \end{aligned}$$

and $l^* A'$ contains a homogeneous arithmetic progression of length L. (A homogeneous arithmetic progression has the form $\{(L_0+1)d, (L_0+2)d, \ldots, (L_0+L)d\}$.)

Recently, Szemerédi and Vu [15] showed that one can guarantee the existence of a (not necessarily homogeneous) arithmetic progression of comparable length under a weaker (and optimal) assumption that $|A'| \ge C\sqrt{n}$, where C is a sufficiently large constant. It is an interesting problem to prove (or disprove) the common strengthening of these two results.

We will apply Theorem 2.1 in conjunction with the lemma below, which allows us to refine an arithmetic progression so that is has relatively small common difference, all while increasing the number of terms compared to the original arithmetic progression.

Lemma 2.2. *Let $A' \subset \{1, 2, \ldots, n\}$ and let $\mathcal{P} \subset S_{A'}$ be an arithmetic progression with length $L = \frac{n}{\gamma}$, where $0 < \gamma < \frac{1}{2}$ is a constant, and with common difference d such that each element of $\mathcal{P}$ is congruent to* $0 \mod d$*. Assume that there exist $d-1$ elements $\{a_1, a_2, \ldots, a_{d-1}\}$ of $\{1, 2, \ldots, n\} \setminus A'$ such that $a_i \equiv r \mod d$ for each i, where r is an integer satisfying $1 \le r \le d-1$. Then, the set $\mathcal{P} + S_{\{a_1,a_2,\ldots,a_{d-1}\}} \subset S_{A'\cup\{a_1,a_2,\ldots,a_{d-1}\}}$ contains an arithmetic* *progression $\mathcal{P}'$ with common difference $d' := \gcd(r, d)$ of length at least $(1-\gamma)\left(\frac{d}{d'}\right)L > L$ such that each element of $\mathcal{P}'$ is congruent to $0 \mod d'$.*

Note that the reason for the hypothesis $0 < \gamma < \frac{1}{2}$ is so that $(1-\gamma)\frac{d}{d'} > 1$ (since $d/d' \geq 2$).

Proof. Consider the sequence of arithmetic progressions

$$\mathcal{P}_k := \begin{cases} \mathcal{P} & \text{if } k = 0, \\ \mathcal{P} + \sum_{i=1}^{k} a_i & \text{if } 1 \leq k \leq d-1. \end{cases}$$

Let p_0 be the smallest element in $\mathcal{P}$. Then the largest element in $\mathcal{P}_0$ is at least $p_0 + Ld$, while the smallest element in $\mathcal{P}_{d-1}$ is at most $p_0 + (d-1)n$. Note that in the range

$$I := [p_0 + (d-1)n,\ p_0 + Ld],$$

every integer that is congruent to $kr \mod d$ is contained in $\mathcal{P}_k$. Thus, inside of I, every integer that is congruent to $0 \mod d'$, where $d' := \gcd(r, d)$, is contained in some $\mathcal{P}_k$. Thus, $\bigcup_{k=0}^{d-1} \mathcal{P}_k$, which is a subset of $\mathcal{P} + S_{\{a_1, a_2, \ldots, a_{d-1}\}}$, contains an arithmetic progression $\mathcal{P}'$ with common difference d' and with length at least

$$\frac{p_0 + Ld - (p_0 + (d-1)n)}{d'} \geq (L-n)\frac{d}{d'} = (1-\gamma)\frac{d}{d'}L + (\gamma L - n)\frac{d}{d'}.$$

By assumption $(1-\gamma)\frac{d}{d'} > 1$ and $\gamma L - n \geq 0$, and by construction, every element of $\mathcal{P}'$ is congruent to $0 \mod d'$. ☐

3. Proof of the main results

3.1. **Proof Theorem 1.5.** We may restate Theorem 1.5 as follows:

Theorem 3.1. *For any constant $c > 0$, there exists a constant $C = C(c) > 0$ such that the following holds for all $\epsilon > 0$ and all integers m and n satisfying*

$$cn(\log n)^{1+\epsilon} \leq m \leq \frac{n^2}{9\log^2 n},$$

where we assume that n is sufficiently large with respect to ϵ and c. If $A \subset \{1, 2, \ldots, n\}$ has cardinality

$$|A| \geq \frac{n}{\operatorname{snd}(m)} + \frac{Cn}{(\log n)^{1+\epsilon}} = (1 + o(1))\frac{n}{\operatorname{snd}(m)},$$

then m can be represented as a sum of distinct elements of A.

Proof. Let $C' := \frac{7\cdot 10^4}{c}$, and let $C := C' + 1$. Let $A' \subset A$ such that $|A'| = \frac{C'n}{(\log n)^{1+\epsilon}}$. By Theorem 2.1, we have that there is an arithmetic progression $\mathcal{P}$ of length $L = 5n \leq \frac{10^{-4}|A'|^2}{\log n}$ and common difference d such that each element in $\mathcal{P}$ is congruent to $0 \mod d$ and such that $\mathcal{P} \subset \ell^* A' \subset S_{A'}$, where $\ell \leq 8496L/|A'| \leq \frac{5c}{7}(\log n)^{1+\epsilon}$. Also, we have that $d \leq 4828n/|A'| \leq \frac{c}{7}(\log n)^{1+\epsilon}$. Now consider the following process.

Step 0: Set $A'_0 := A'$, set $B_0 := A \setminus A'_0$, set $\mathcal{P}_0 := \mathcal{P}$, and set $d_0 := d$.

Step i: (a) Look at the elements of B_i modulo d_i. If for each $1 \leq r \leq d_i - 1$ there are at most $d_i - 2$ elements in B_i congruent to $r \mod d_i$, then STOP. Otherwise, go to (b).

(b) Let $1 \le r \le d_i - 1$ be an integer such that there are at least $d_i - 1$ elements of B_i congruent to $r \mod d_i$ and such that $\gcd(r, d_i)$ is as small as possible. Call this set of $d_i - 1$ elements $B_i' \subset B_i$. By Lemma 2.2 (with $\gamma = 1/5$), we know that $\mathcal{P}_i + S_{B_i'} \subset S_{A_i' \cup B_i'}$ contains an arithmetic progression $\mathcal{P}_{i+1}$ of length at least L and with common difference $d_{i+1} := \gcd(r, d_i)$. Set $A_{i+1}' := A_i' \cup B_i'$ and set $B_{i+1} := A \setminus A_{i+1}'$. Now go to step $i + 1$.

Note that $d_{i+1} \le d_i/2$, and thus the algorithm can take at most $\log_2 d = O(\log n)$ steps. Thus, at the final step, say t, we have

$$\begin{aligned} |B_t| &\ge |A| - |A'| - d(1 + 1/2 + 1/4 + \cdots + 1/2^{t-1}) \\ &\ge \frac{n}{\operatorname{snd}(m)} + \frac{Cn}{(\log n)^{1+\epsilon}} - \frac{C'n}{(\log n)^{1+\epsilon}} - 2 \cdot \frac{c}{7}(\log n)^{1+\epsilon} \\ &\ge \frac{n}{\operatorname{snd}(m)} + \frac{3}{4}\left(\frac{n}{(\log n)^{1+\epsilon}}\right), \end{aligned}$$

for sufficiently large n.

Also note that at the final step t, at most $(d-1)^2 \le \frac{c^2}{49}(\log n)^{2+2\epsilon}$ elements of B_t are not congruent to $0 \mod d_t$. Thus, B_t contains at least

$$|B_t| - \frac{c^2}{49}(\log n)^{2+2\epsilon} \ge \frac{n}{\operatorname{snd}(m)} + \frac{1}{2}\left(\frac{n}{(\log n)^{1+\epsilon}}\right) > \frac{n}{\operatorname{snd}(m)}$$

elements that are congruent to $0 \mod d_t$ (again, assuming that n is sufficiently large). But $\{1, 2, \ldots, n\}$ contains only n/d_t elements congruent to $0 \mod d_t$, and so we must have that $d_t < \operatorname{snd}(m)$. This key fact implies, by the definition of $\operatorname{snd}(m)$, that d_t divides m.

Now, let $\{b_1, b_2, \ldots, b_{k_0}\}$ be elements of B_t congruent to $0 \mod d_t$, where $k_0 = \left\lfloor \frac{n}{\operatorname{snd}(m)} \right\rfloor$. We will "grow" the arithmetic progression so that it is long enough to contain m. Recall that $\mathcal{P}_t$ is the final arithmetic progression constructed by the process above, and consider the sequence of arithmetic progressions

$$\mathcal{Q}_k := \begin{cases} \mathcal{P}_t & \text{if } k = 0, \\ \mathcal{P}_t + \sum_{i=1}^k b_i & \text{if } 1 \le k \le k_0. \end{cases}$$

Note that $\mathcal{Q}_{k-1}$ overlaps with $\mathcal{Q}_k$ for all $1 \le k \le k_0$, since $\mathcal{P}_t$ has length greater than n and since all elements in $\mathcal{P}_t$ and in $\{b_1, b_2, \ldots, b_{k_0}\}$ are congruent to $0 \mod d_t$. Thus, S_A contains an arithmetic progression $\mathcal{Q} = \bigcup_{k=0}^{k_0} \mathcal{Q}_k$ with common difference $d_t < \operatorname{snd}(m)$ and with each element of the arithmetic progression congruent to $0 \mod d_t$.

The largest element in $\mathcal{Q}$ is at least

$$\sum_{i=1}^{k_0+1} i \ge \frac{n^2}{2\operatorname{snd}(m)^2} \ge \frac{n^2}{9\log^2 n},$$

using the fact that (by the prime number theorem) $\operatorname{snd}(m) \le (2 + o(1))\log n$. On the other hand, the smallest element in $\mathcal{Q}$ (which is the same as the smallest element in $\mathcal{P}_t$) is at most

$$n(\ell + d(1 + 1/2 + \cdots + 1/2^{t-1})) \le n(\ell + 2d) \le c(\log n)^{1+\epsilon}$$

By assumption, we have $cn(\log n)^{1+\epsilon} \le m \le \frac{n^2}{9\log^2 n}$, and so we see that $\mathcal{Q}$ contains m, completing the proof. $\square$

3.2. **Proof of Theorem 1.6.** We may restate Theorem 1.6 as follows:

Theorem 3.2. *For any constant $c > 0$, there exists a constant $C = C(c) > 0$ such that the following holds for all constants $0 < \epsilon_1 < \epsilon_2$, for all $\delta > 0$, and for all integers m and n satisfying*

$$cn(\log n)^{1+\epsilon_2} \leq m \leq \frac{\delta^2 n^2}{8(\log n)^{2+2\epsilon_1}},$$

where we assume that n is sufficiently large with respect to c, ϵ_1, ϵ_2, and δ. If $A \subset \{1, 2, \ldots, n\}$ has cardinality

$$|A| \geq \frac{\delta n}{(\log n)^{1+\epsilon_1}}$$

and if m cannot be represented as a sum of distinct elements in A, then A contains at least

$$\frac{\delta n}{(\log n)^{1+\epsilon_1}} - \frac{Cn}{(\log n)^{1+\epsilon_2}} = (1 - o(1))\,|A|$$

elements that are congruent to $0 \mod d$, where d is an integer that does not divide m.

One can prove Theorem 3.2 using the same proof as for Theorem 3.1 (with a few small changes). Here we only sketch the proof. For a set A satisfying the conditions of theorem 3.2, let $A' \subset A$ be such that $|A'| = \frac{Cn}{3(\log n)^{1+\epsilon_2}}$. By Theorem 2.1 we can find a long arithmetic progression $\mathcal{P} \subset S_{A'}$ and refine it by Lemma 2.2. If the refining process ends after t steps then we have an arithmetic progression $\mathcal{P}_t$ with common difference d_t. The set $B_t \subset A$ will contain at least

$$\frac{\delta n}{(\log n)^{1+\epsilon_1}} - \frac{Cn}{(\log n)^{1+\epsilon_2}} \tag{2}$$

elements that are congruent to $0 \mod d_t$. After "growing" $\mathcal{P}_t$ using these elements we have a long arithmetic progression $\mathcal{Q}$ with elements that are congruent to $0 \mod d_t$, with common difference d_t, and containing elements both smaller and larger than m. If m is congruent to $0 \mod d_t$ then $m \in \mathcal{Q} \subset S_A$, a contradiction; thus, d_t does not divide m and the theorem is proved.

References

1. N. Alon, *Subset sums*, J. Number Theory **27** (1987), no. 2, 196–205. MR MR909836 (88k:11020)
2. N. Alon and G. Freiman, *On sums of subsets of a set of integers*, Combinatorica **8** (1988), no. 4, 297–306. MR MR981887 (90e:11028)
3. J. Bourgain, *On arithmetic progressions in sums of sets of integers*, A tribute to Paul Erdős, Cambridge Univ. Press, Cambridge, 1990, pp. 105–109. MR MR1117007 (92e:11011)
4. Paul Erdős, *Some problems and results on combinatorial number theory*, Graph theory and its applications: East and West (Jinan, 1986), Ann. New York Acad. Sci., vol. 576, New York Acad. Sci., New York, 1989, pp. 132–145. MR MR1110810 (92g:11011)
5. Paul Erdős and Gregory Freiman, *On two additive problems*, J. Number Theory **34** (1990), no. 1, 1–12. MR MR1039762 (91c:11011)
6. G. A. Freiman, H. Halberstam, and I. Z. Ruzsa, *Integer sum sets containing long arithmetic progressions*, J. London Math. Soc. (2) **46** (1992), no. 2, 193–201. MR MR1182477 (93j:11008)
7. E. Lipkin, *On representation of rth powers by subset sums*, Acta Arith. **52** (1989), no. 4, 353–365. MR MR1030087 (91c:11055)
8. Leo Moser and Peter Scherk, *Advanced Problems and Solutions: Solutions: 4466*, Amer. Math. Monthly **62** (1955), no. 1, 46–47. MR MR1528921
9. Hoi H. Nguyen, Endre Szemerédi, and Van H. Vu, *Subset sums modulo a prime*, Acta Arith. **131** (2008), no. 4, 303–316. MR MR2383688
10. Hoi H. Nguyen and Van H. Vu, *Classification theorems for sumsets modulo a prime, to appear,* J. Combin. Theory Ser. A (2008).

11. K. F. Roth, *On certain sets of integers*, J. London Math. Soc. **28** (1953), 104–109. MR MR0051853 (14,536g)
12. A. Sárközy, *Finite addition theorems. I*, J. Number Theory **32** (1989), no. 1, 114–130. MR MR1002119 (91f:11007)
13. ______, *Finite addition theorems. III*, Groupe de Travail en Théorie Analytique et Élémentaire des Nombres, 1989–1990, Publ. Math. Orsay, vol. 92, Univ. Paris XI, Orsay, 1992, pp. 105–122. MR MR1192813 (93i:11010)
14. ______, *Finite addition theorems. II*, J. Number Theory **48** (1994), no. 2, 197–218. MR MR1285539 (95k:11010)
15. E. Szemerédi and V. Vu, *Long arithmetic progressions in sumsets: thresholds and bounds*, J. Amer. Math. Soc. **19** (2006), no. 1, 119–169 (electronic). MR MR2169044 (2006j:11015)
16. E. Szemerédi and V. H. Vu, *Long arithmetic progressions in sum-sets and the number of x-sum-free sets*, Proc. London Math. Soc. (3) **90** (2005), no. 2, 273–296. MR MR2142128 (2006a:11011)
17. ______, *Finite and infinite arithmetic progressions in sumsets*, Ann. of Math. (2) **163** (2006), no. 1, 1–35. MR MR2195131 (2006j:11014)
18. Van H. Vu, *A structural approach to subset-sum problems*, Building Bridges: Between Mathematics and Computer Science (Martin Grötschel and Gyula O.H. Katona, eds.), Bolyai Society Math. Studies, vol. 19, Springer, 2008, pp. 525–545.

Department of Mathematics, Rutgers, Piscataway, NJ 08854
E-mail address: `linhtran@math.rutgers.edu`

Department of Mathematics, Rutgers, Piscataway, NJ 08854
E-mail address: `vanvu@math.rutgers.edu`

Department of Mathematics, Rutgers, Piscataway, NJ 08854
E-mail address: `matchett@math.rutgers.edu`