COUNTING GENERALIZED SUM-FREE SETS

NEIL J. CALKIN AND JAN MCDONALD THOMSON

Abstract. We show that the number of subsets of $\{1, 2, ..., n\}$ with no solution to $x_1 + x_2 + ... + x_k = y_1 + y_2 + .. + y_l$ for $k \geq 4l - 1$ is at most $c2^{\theta n}$ where $\theta = \frac{k-l}{k}$.

1. Introduction

A set S of positive integers is *sum-free* if $x + y = z$ has no solution in S. Similarly, a set S of positive integers is (k, l) sum-free if $x_1 + x_2 + ... + x_k = y_1 + y_2 + ... + y_l$ has no solution in S. Cameron and Erdös have shown [5] that the number of sum-free sets contained in $\{\frac{1}{3}n, ..., n\}$ is $c2^{\frac{n}{2}}$, and Alon [1], Calkin [3] and Erdös and Granville (personal communication) have independently shown that the number of sum-free sets contained in $\{1, 2, ..., n\}$ is $o\left(2^{n(\frac{1}{2}+\varepsilon)}\right)$ for every $\varepsilon > 0$. Calkin and Taylor [4] have shown that the number of $(k, 1)$ sum-free sets in $\{1, 2, ..., n\}$ is $c2^{\theta n}$ for $k \geq 3$ and $\theta = \frac{k-1}{k}$. Bilu [2] has shown that the number of $(4, 3)$ sum-free sets, and consequently, the number of $(k + 1, k)$ sum-free sets is at most $c2^{(n+1)/2}$ for $k \geq 3$. We extend the results of Calkin and Taylor in a different direction by showing that for $k \geq 4l - 1$, the number of (k, l) sum-free sets in $\{1, 2, ..., n\}$ is at most $c2^{\theta n}$ where $\theta = \frac{k-l}{k}$. Note that the number of (k, l) sum-free sets in $\{1, 2, ..., n\}$ is at least $c2^{\theta n}$ because any subset of $\{\frac{l}{k}n, ..., n\}$ is (k, l) sum-free, and there are $2^{n-\frac{l}{k}n} = 2^{\theta n}$ such subsets. Our main result is the following theorem.

Theorem 1. *For fixed k, l with $k \geq 4l - 1$, let $\theta = \frac{k-l}{k}$. There exists a constant c depending on k such that the number of subsets of $\{1, 2, ..., n\}$ containing no solution to $x_1 + x_2 + ... + x_k = y_1 + y_2 + ... + y_l$ is at most $c2^{\theta n}$.*

1991 *Mathematics Subject Classification.* 11B13,05A16.

2. Preliminaries

We will denote by [a,b] the set [a,b]$\cap\mathbb{Z}$. We will fix a large constant q, depending on k and l, and consider separately those sets with respectively few (less than εq) and many (at least εq) elements in $\{n-q,...,n\}$ for some fixed ε. We will see that there are relatively few sets of the former type and that those of the latter will contain long arithmetic progressions which we will make use of in our proof.

Specifically we define:

$$x_1 + x_2 + ... + x_k = y_1 + y_2 + ... + y_l \tag{1}$$

$(*)$-free to mean a set having no solution to (1)

$$\begin{aligned}
f_n &= \text{ the number of } (*)\text{-free sets in } \{1,2,...,n\} \\
g_n &= \text{ the number of } (*)\text{-free sets in } \{1,2,...,n\} \text{ which contain less} \\
&\quad \text{than } \varepsilon q \text{ elements greater than } n-q. \\
h_{n,m} &= \text{ the number of } (*)\text{-free sets in } \{1,2,...,n\} \text{ which contain at} \\
&\quad \text{least } \varepsilon q \text{ elements greater than } n-q \text{ and have least element } m. \\
h_n &= \sum_{m=1}^{n} h_{n,m} = \text{ the number of } (*)\text{-free sets in } \{1,2,...,n\} \text{ which} \\
&\quad \text{contain at least } \varepsilon q \text{ elements greater than } n-q. \\
d &= \max\left(\left\lceil \frac{1}{2\theta-1} \right\rceil, 8l\right)
\end{aligned}$$

We will make use of the following theorem of Szemerédi [6]:

Theorem (Szemerédi)**.** *Given d, $\varepsilon > 0$ there exists q_o so that if $q > q_o$ then any subset S of $\{1,2,...,q\}$ with $|S| > \varepsilon q$ contains an arithmetic progression of length greater than $2d$.*

Thus we may choose $\varepsilon > 0$, and q sufficiently large so that

(1) If a subset S of $\{n-q+1,...,n\}$ is such that $|S| > \varepsilon q$ then there is an arithmetic progression of length at least $2d+1$.
(2) $\binom{q}{\varepsilon q}\varepsilon q < 2^{\theta q - 1}$.

3. Counting sets with fewer than εq elements in $[n-q+1,n]$

We first find a bound on g_n, i.e., how many $(*)$- free sets there are in $\{1,2,...,n\}$ which contain less than εq elements greater than $n-q$.

The number of ways to get fewer than εq elements in $[n-q+1, n]$ is

$$\sum_{i=1}^{\varepsilon q} \binom{q}{\varepsilon q - i} \leq \varepsilon q \binom{q}{\varepsilon q} < 2^{\theta q - 1}$$

If we now multiply this by the number of $(*)$-free sets in $\{1, 2, ..., n-q\}$, we obtain an upper bound for g_n of

$$2^{\theta q - 1} f_{n-q}.$$

Next, we will bound h_n and the result will follow by induction.

4. Counting sets with more than εq elements in $[n-q+1, n]$

In this section we will find bounds on $h_{n,m}$ for various values of the least element, m, of a set. Let us first consider the case where $m > \frac{l}{k}n$. The number of (k, l)-sum-free sets with $m > \frac{l}{k}n$ does not exceed the number of subsets of $[\frac{l}{k}n + 1, n]$, which is less than $2^{\theta n}$.

We are left with the case where $m \leq \frac{l}{k}n$. In this case we will use the arithmetic progression of length $2d+1$ in $[n-q+1, n]$ that we are guaranteed by Szemerédi. Call this progression $t-da, t-(d-1)a, ..., t+da$. To make use of this progression, we will need the following lemma due to Calkin [3].

Lemma 2. *Given an arithmetic progression $b-da, b-(d-1)a, ..., b+da$, the number of subsets of $\{0, 1, ..., b-1\}$ having no pairs x, y such that $x+y$ is an element of the progression, is at most*

$$2^{\frac{d+1}{2d}(b+a(2d+1))+1}.$$

Clearly, $h_{n,m}$ is less than the number of subsets of $\{m, ..., n\}$ having no solution to $x_1 + x_2 + (k-2)m = y_1 + y_2 + ... + y_l$. Making the substitutions $z_1 = x_1 + m$ and $z_2 = x_2 + m$, we have $h_{n,m}$ less than the number of subsets A of $\{0, ...n-m\}$ such that for any $z_1, z_2 \in A$ and $y_1, ..., y_l \in A + m$ we have $z_1 + z_2 \neq y_1 + y_2 + ... + y_l$. Various substitutions for $y_1$through y_l will give us the desired result.

We will split the interval $\{1, ..., \frac{l}{k}n\}$ into overlapping intervals $I_p = [\alpha_p n, \beta_p n]$, $p = 1, ..., l$ and $J_p = [\alpha'_p n, \beta'_p n]$ for $p = 1, ..., l-1$ where

$$\alpha_p = \frac{p\frac{t}{n} - 1}{k - l + p - 1} \qquad \beta_p = \frac{kp - 2l - \frac{1}{d}kp}{k(k-l+p-2) - \frac{1}{d}k(k-l+p)}$$

$$\alpha'_p = \frac{k(p-1) + 2l + \frac{1}{d}(p+1)k}{k(k-l+p+1) + \frac{1}{d}k(k-l+p+1)} \qquad \beta'_p = \frac{(p+1)\frac{t}{n} - 1}{k-l+p}$$

and look at the cases $m \in I_p$, $p = 1, ..., l$ $m \in J_p$, $p = 1, .., l-1$ separately.

5. Sets with $m \in I_p$ for $p = 1, ..., l$

We know that $h_{n,m}$ is less than the number of subsets A of $\{0, ..., n-m\}$ such that for any $z_1, z_2 \in A$ and $y_1, ..., y_l \in A+m$ we have $z_1+z_2 \neq y_1+y_2+...+y_l$. Therefore, letting $y_1 \in \{t-da, t-(d-1)a, ..., t+da\}$, $y_2,\ y_3, ...,\ y_p = t$, and $\ y_{p+1}, ...,\ y_l = m$, the number $h_{n,m}$ is still less than the number of subsets of $\{0, ..., n-m\}$ with no solution to

$$z_1 + z_2 = (pt - (k-l+p)m) + ia \qquad \text{for } i = -d, ..., d \tag{2}$$

By Lemma 1 there are at most $2^{\frac{d+1}{2d}(pt-(k-l+p)m+a(2d+1))+1}$ subsets of $\{0, 1, ..., pt-(k-l+p)m-1\}$ with no solution to (2). Note that we know $pt-(k-l+p)m \leq n-m$. If we now allow any subset of the remaining elements in $\{pt-(k-l+p)m, ..., n-m\}$ we have

$$\begin{aligned} h_{n,m} &\leq 2^{\frac{d+1}{2d}(pt-(k-l+p)m+a(2d+1))+1} 2^{(n-m)-(pt-(k-l+p)m)+1} \\ &= 2^{\gamma_p n + \delta_p m + \frac{d+1}{2d}(a(2d+1))+2+(n-t)p(1-\frac{d+1}{2d})} \end{aligned}$$

where $\gamma_p = 1 + \frac{p}{2}\left(\frac{1}{d} - 1\right)$ and $\delta_p = \frac{1}{2}(k-l+p)\left(1-\frac{1}{d}\right) - 1$. Our choice of β_p, γ_p, and δ_p insure that $\gamma_p + \delta_p\beta_p = \theta$. Also note that $(n-t) < q$, so for $\epsilon_p = ql + 2$

$$h_{n,m} < 2^{\gamma_p n + \delta_p m + \epsilon_p} \tag{3}$$

Therefore

$$\begin{aligned} \sum_{m \in I_p} h_{n,m} &< \sum_{m \in I_p} 2^{\gamma_p n + \delta_p m + \epsilon_p} \\ &= 2^{\gamma_p n + \epsilon_p} \sum_{m \in I_p} 2^{\delta_p m} \\ &\leq 2^{\gamma_p n + \epsilon_p} \left(\frac{2^{(\beta_p n + 1)\delta_p} - 2^{\alpha_p \delta_p n}}{2^{\delta_p} - 1} \right) \\ &= 2^{(\gamma_p + \delta_p \beta_p) n + \epsilon_p} \left(\frac{2^{\delta_p} - 2^{-\delta_p(\beta_p - \alpha_p) n}}{2^{\delta_p} - 1} \right), \end{aligned}$$

so there exists a c_p for which

$$\sum_{m \in I_p} h_{n,m} < c_p 2^{\theta n} \tag{4}$$

6. Sets with $m \in J_p$ for $p = 1, ..., l-1$

Again we will use the fact that $h_{n,m}$ is less than the number of sets A of $\{0, ..., n-m\}$ such that for any $z_1, z_2 \in A$ and $y_1, ..., y_l \in A+m$ we have $z_1 + z_2 \neq y_1 + y_2 + ... + y_l$. This time we will use the substitutions $y_1 \in \{t-da, t-(d-1)a, ..., t+da\}$, $y_2, y_3, ..., y_{p+1} = t$, and $y_{p+2}, ..., y_l = m$. We are now looking at sets with no solution to

$$z_1 + z_2 = ((p+1)t - (k-l+p+1)m) + ia \qquad i = -d, ..., d. \tag{5}$$

Using Lemma 1 for an upper bound on the number of subsets of $\{0, 1, ..., (p+1)t - (k-l+p+1)m - 1\}$ with no solution to (5) we have

$$\begin{aligned} h_{n,m} &\leq 2^{\frac{d+1}{2d}((p+1)t-(k-l+p+1)m+a(2d+1))+1} \\ &= 2^{\gamma'_p n - \delta'_p m + \frac{d+1}{2d}(a(2d+1)-(p+1)(n-t))+1} \end{aligned}$$

where $\gamma'_p = \frac{d+1}{2d}(p+1)$ and $\delta'_p = \frac{d+1}{2d}(k-l+p+1)$. Note that we have not counted any subsets of $\{(p+1)t - (k-l+p+1)m, ..., n-m\}$ because this set is empty for $m \in J_p$. To be precise, this set may contain one element; this happens when $m = \beta_p \prime n$. Here, our choice of α'_p, γ'_p, and δ'_p insure that $\gamma'_p - \delta'_p \alpha'_p = \theta$. For $\epsilon'_p = q$, we have $h_{n,m} < 2^{\gamma'_p n - \delta'_p m + \epsilon'_p}$. Therefore

$$\begin{aligned} \sum_{m \in J_p} h_{n,m} &< \sum_{m \in J_p} 2^{\gamma'_p n - \delta'_p m + \epsilon'_p} \\ &= 2^{\gamma'_p n + \epsilon'_p} \sum_{m \in J_p} 2^{-\delta'_p m} \\ &\leq 2^{\gamma'_p n + \epsilon'_p} \left(\frac{2^{-\alpha'_p \delta'_p n} - 2^{-(\beta'_p n + 1)\delta'_p}}{1 - 2^{-\delta'_p}} \right) \\ &= 2^{(\gamma'_p - \delta'_p \alpha'_p)n + \epsilon'_p} \left(\frac{1 - 2^{-\delta'_p((\beta'_p - \alpha'_p)n+1)}}{1 - 2^{-\delta'_p}} \right). \end{aligned}$$

Again, there is a c'_p for which

$$\sum_{m \in J_p} h_{n,m} < c'_p 2^{\theta n} \tag{6}$$

7. Checking that I_p and J_p cover $[0, \frac{l}{k}n]$

In this section we prove that

$$[0, ..., \frac{l}{k}n] \subseteq I_1 \cup J_1 \cup I_2 \cup J_2 \cup ... \cup I_{l-1} \cup J_{l-1} \cup I_l.$$

It suffices to show that:

$$\begin{aligned} \alpha_1 &\le 0, & \beta_l &= \frac{l}{k}, \\ \alpha_{p+1} &= \beta_{p'} & &(1 \le p \le l-1), \\ \alpha_{p'} &\le \beta_p & &(1 \le p \le l-1). \end{aligned}$$

Here the first two are obvious. It remains to prove $\alpha_{p'} \le \beta_p$, which is equivalent to showing

$$\frac{k(p-1) + 2l + \frac{1}{d}k(p+1)}{k(k-l+p+1) + \frac{1}{d}k(k-l+p+1)} \le \frac{kp - 2l - \frac{1}{d}kp}{k(k-l+p-2) - \frac{1}{d}k(k-l+p)}$$

for $1 \le p \le l-1$. Since $d > 3$, both denominators are positive and we can cross multiply and get the following equivalent expression

$$\frac{1}{d}(k-l)\left(2k - 2 - \frac{1}{d}k\right) \le (k-l)\left(k - 4l + 4p - 2\right).$$

The right hand side of the above expression is smallest when $p = 1$ so we must have the following

$$k \ge \frac{4l - 2 - \frac{2}{d}}{\left(1 - \frac{1}{d}\right)^2} \qquad \text{since } k > l.$$

Since we know $k \ge 4l - 1$, if we can show that $\frac{4l-2-\frac{2}{d}}{\left(1-\frac{1}{d}\right)^2} \le 4l - 1$, then we are done. Notice

$$\frac{4l - 2 - \frac{2}{d}}{\left(1 - \frac{1}{d}\right)^2} \le \frac{4l - 2}{\left(1 - \frac{1}{d}\right)^2} \le 4l - 1$$

as soon as

$$d \ge d_0(l) := \frac{1}{1 - \sqrt{\frac{4l-2}{4l-1}}}$$

so $d \ge 8l$ will suffice.

8. Putting it all together

Now, since we have bounded $h_{n,m}$ for all values of m, we are in a position to bound h_n.

$$
\begin{aligned}
h_n &= \sum_{m=1}^{n} h_{n,m} \\
&= \sum_{m \le \frac{l}{k}n} h_{n,m} + \sum_{m > \frac{l}{k}n} h_{n,m} \\
&= \sum_{p=1}^{l} \sum_{m \in I_p} h_{n,m} + \sum_{p=1}^{l-1} \sum_{m \in J_p} h_{n,m} + \sum_{m > \frac{l}{k}n} h_{n,m} \\
&\le \sum_{p=1}^{l} c_p 2^{\theta n} + \sum_{p=1}^{l-1} c_p' 2^{\theta n} + 2^{\theta n} \\
&\le c_h 2^{\theta n}
\end{aligned}
$$

for some c_h depending upon k and l. Note that we do not claim these bounds to be tight.

We now have bounds on g_n and h_n and are able to bound f_n, or the total number of $(*)$-free sets in $\{1, 2, ..., n\}$. We will show that we can choose c independent of n.

Proof of Theorem 1: Our proof will proceed by induction. Let us first notice that the work we have done so far insures that

$$
\begin{aligned}
f_n &= g_n + h_n \\
&\le 2^{\theta q - 1} f_{n-q} + c_h 2^{\theta n}
\end{aligned}
\tag{7}
$$

Observe that $f_n < 2^q 2^{\theta n}$ for $1 \le n \le q$: hence if we set $c > \max(2c_h, 2^q)$ then we obtain

$$
\begin{aligned}
f_n &< \left(\frac{c}{2} + c_h\right) 2^{\theta n} \\
&< c 2^{\theta n}
\end{aligned}
$$

which is our desired result. □

We are grateful to a referee for pointing out that the assumption $k \ge 4l - 1$ can be relaxed to $k \ge 4l - \gcd(k, l)$: indeed, if the assertion holds for a pair (k, l), then it also holds for (dk, dl).

9. Conjectures

We believe that our result is still true for all $k \ge 2l$ but that significantly different techniques must be used to prove it. Note that our intervals I_p and J_p are no longer guaranteed to overlap if $k < 4l-1$. For

$k = 2l$ we have sum-free sets as a special case (l=1), so we also make the weaker conjecture that the number of (k, l) sum-free sets contained in $\{1, 2, ..., n\}$ is $c2^{\theta n}$ for $\theta = \frac{k-l}{k}$ and $k > 2l$.

It is also natural to ask whether an asymptotic result is true: that is, does there exist a constant c such that

$$f_n \sim c2^{\theta n}$$

as $n \to \infty$? Computational evidence for sum-free sets suggests that this is not the case: there seems to be a parity effect so that for even n and for odd n there are different asymptotic results. This reflects the fact that there are two distinct large sum-free sets: the odd numbers, and the numbers in $[n/2, n]$. Since this is not the case for (k, l) sum-free sets, we conjecture that there is such a constant.

References

[1] Noga Alon. Independent sets in regular graphs and sum-free subsets of finite groups. *Israel J. Math.,* 73:247-256, 1991.

[2] Yuri Bilu. Sum-free sets and related sets, preprint.

[3] Neil J. Calkin. On the number of sum-free sets. *Bull. London Math. Soc.,* 22:141-144,1990.

[4] Neil J. Calkin and Angela C. Taylor. Counting sets of integers, no k of which sum to another. *Journal of Number Theory,* pages 323-327, 1996.

[5] P. J. Cameron and P. Erdös. On the number of sets of integers with various properties. In Richard A. Mollin, editor, *Number Theory,* pages 61-79. Walter de Gruyter, 1990.

[6] E. Szemerédi. On sets of integers containing no k elements in arithmetic progression. *Acta. Arith.,* 27:199-245, 1975.

E-mail address: `calkin@math.gatech.edu, thomson@math.gatech.edu`

School of Mathematics, Georgia Istitute of Technology, Atlanta, GA 30332-0160