On the number of sum-free sets.

Neil J. Calkin
Department of Mathematics,
Carnegie Mellon University,
Pittsburgh, PA

Abstract

Cameron and Erdös have considered the question: how many sum-free sets are contained in the first n integers; they have shown that the number of sum-free sets contained within the integers $\{\frac{n}{3}, \frac{n}{3}+1, \ldots, n\}$ is $c.2^{\frac{n}{2}}$. We prove that the number of sets contained within $\{1, 2, \ldots, n\}$ is $o(2^{n(1/2+\varepsilon)})$ for every $\varepsilon > 0$.

1 Introduction

A set S is said to be **sum-free** if the equation $x+y=z$ has no solutions within S. Cameron [3],[4],[5] has asked many questions regarding sum-free sets; in particular he has conjectured that the Hausdorff dimension of $\mathcal{S}$, the set of all sum-free sets of positive integers, is equal to 1/2; he further observed that this would follow immediately if the number of sum-free sets contained in the set $\{1, 2, \ldots, n\}$ is $o(2^{n(1/2+\varepsilon)})$ for every $\varepsilon > 0$. Cameron and Erdös (personal communication) have shown that the number of sum-free sets contained in $\{\frac{n}{3}, \frac{n}{3}+1, \ldots, n\}$ is $c.2^{\frac{n}{2}}$, and Calkin [1],[2] has shown that the Hausdorff dimension of $\mathcal{S}$ is at most .599. In this paper, we show that the number of sum-free sets in $\{1, 2, \ldots, n\}$ is indeed $o(2^{n(1/2+\varepsilon)})$ for every $\varepsilon > 0$, and immediately deduce that the Hausdorff dimension of $\mathcal{S}$ is 1/2.

Erös and Granville (personal communication) have independently proven the same result; their method also uses a theorem due to Szemerédi, in graph theory.

2 The Main Theorem

Theorem 1 *For every $\varepsilon > 0$, the number of sum-free sets contained in the set $\{1, 2, \ldots, n\}$ is $o(2^{n(1/2+\varepsilon)})$.*

We shall defer the proof for a moment; first we shall state several results required for the proof. The first is Szemeredi's celebrated theorem on arithmetic progressions.

Theorem 2 (Szemeredi) *There exists a function $g_k(n)$ such that $g_k(n) = o(n)$ and every subset of size $g_k(n)$ from the integers $\{1, 2, \dots, n\}$ contains an arithmetic progression of length at least k.*

Proof. See Szemeredi [6].

Lemma 1 *For every $\varepsilon > 0$, the number of subsets of size at most $f(n)$ of $\{1, 2, \dots, n\}$ is $o(2^{\varepsilon n})$ whenever $f(n)$ is $o(n)$.*

Proof. The number of such subsets is at most

$$f(n)\binom{n}{f(n)} < f(n)\frac{n^{f(n)}}{f(n)!} < f(n)(\frac{ne}{f(n)})^{f(n)}$$

$$< e^{f(n)\log n - f(n)\log f(n) + f(n) + \log f(n)}$$

$$= e^{n(\frac{f(n)}{n}\log n - \frac{f(n)}{n}\log f(n) + \frac{f(n)}{n} + \frac{\log f(n)}{n})}$$

$$= e^{n(-\frac{f(n)}{n}\log\frac{f(n)}{n} + \frac{f(n)}{n} + \frac{\log f(n)}{n})}$$

and since $f(n) = o(n)$ and $x \log x \to 0$ as $x \to 0$, this is $o(2^{\varepsilon n})$ for every $\varepsilon > 0$.

Lemma 2 *The number of binary sequences of length m without any pair of 1's at distance exactly 1,3,5,7,. . . ,$2k - 1$, is at most $2^{\frac{k+1}{2k}(m+2k)}$.*

Proof. The number of sequences of length $2k$ without pairs of 1's at an odd distance is exactly $2^{k+1} - 1$. Thus the number of sequences of length m without pairs of 1's at an odd distance less than $2k$ is at most

$$(2^{k+1} - 1)^{\lceil \frac{m}{2k} \rceil} < (2^{k+1})^{\frac{m}{2k}+1} = 2^{\frac{k+1}{2k}(m+2k)}$$

as required.

Lemma 3 *Given an arithmetic progression $m - kd, m - (k - 1)d, \dots, m, \dots, m + kd$, the number of subsets of $\{1, 2, \dots, m - 1\}$ having no pairs x, y such that $x + y$ is an element of the progression, is at most*

$$2^{\frac{k+1}{2k}(m+d(2k+1))}$$

Proof. Write the elements of $\{1, 2, \dots, m - 1\}$ in the following d sequences;

$$A_1 = 1, m - 1, 1 + d, m - 1 - d, 1 + 2d, m - 1 - 2d, \dots,$$

$$A_2 = 2, m - 2, 2 + d, m - 2 - d, 2 + 2d, m - 2 - 2d, \dots$$

$$\vdots$$

$$A_d = \{d, m - d, 2d, m - 2d, 3d, m - 3d, \dots\}$$

where each sequence has either $\lfloor \frac{m}{d} \rfloor$ or $\lceil \frac{m}{d} \rceil$ elements, and every element of $\{1, 2, \ldots, m\}$ occurs in exactly one such sequence. Then, for any set S which has no pair of elements summing to a member of the arithmetic progression, the characteristic sequence of S is such that when it is written as d binary sequences in the order given by $A_1, \ldots, A_d$, each of these binary sequences has the property that there are no 1's at distance exactly 1,3,5,7,…,$2k-1$. The number of ways of choosing such a set S is thus at most the number of ways of choosing d sequences of length $\frac{m}{d}+1$, without 1's at an odd distance less than $2k$. This is at most

$$2^{\frac{k+1}{2k}(\frac{m}{d}+1+2k)d} = 2^{\frac{k+1}{2k}(m+d(2k+1))}$$

as desired.

We are now in a position to prove the theorem; let $\varepsilon > 0$, and fix $k > \frac{1}{\varepsilon}$; partition the set $\{1, 2, \ldots, n\}$ into $\lceil n^{1/2} \rceil$ disjoint intervals of size as nearly equal as possible, that is either $\lfloor n^{1/2} \rfloor$ or $\lceil n^{1/2} \rceil$. Then every S with at least $\lceil n^{1/2} \rceil g_{2k+1}(\lceil n^{1/2} \rceil)$ elements contains at least $g_{2k+1}(\lceil n^{1/2} \rceil)$ elements in one of these intervals. Let $\mathcal{S}_p$ be the set of sum-free sets which contain at least $g_{2k+1}(\lceil n^{1/2})$ in the pth interval, and fewer than $g_{2k+1}(\lceil n^{1/2} \rceil)$ in every subsequent interval; that is the pth interval is the last one with which S has a large intersection. Then, for every $S \in \mathcal{S}_p$, S contains an arithmetic progression

$$m-kd, m-(k-1)d, \ldots, m, m+d, \ldots, m+kd$$

which lies in the pth interval.

How many possible arithmetic progressions of this form are there in the pth interval? Clearly there are at most $\lceil n^{1/2} \rceil$ choices for m, and at most $\lceil n^{1/2} \rceil$ choices for d.

How many possible sum-free sets in $\{1, 2, \ldots, n\}$ contain the arithmetic progression

$$m-kd, m-(k-1)d, \ldots, m, m+d, \ldots, m+kd?$$

From Lemma 2 and Lemma 3 we see that the number of such sets is at most

$$2^{\frac{k+1}{2k}(m+d(2k+1))} 2^{\lceil n^{1/2} \rceil} \binom{n-p\lfloor n^{1/2} \rfloor}{(\lceil n^{1/2} \rceil - p) g_{2k+1}(\lceil n^{1/2} \rceil)}$$

$$\leq 2^{\frac{k+1}{2k}(m+d(2k+1))} 2^{\lceil n^{1/2} \rceil} \binom{\lceil n^{1/2} \rceil(\lceil n^{1/2} \rceil - p)}{(\lceil n^{1/2} \rceil - p) g_{2k+1}(\lceil n^{1/2} \rceil)}$$

Now m is at most n, and d is at most $\frac{\lceil n^{1/2} \rceil}{2k} < \frac{2n^{1/2}}{2k+1}$, so the product of the first two factors is at most

$$2^{\frac{k+1}{2k}(n+2n^{1/2})} 2^{n^{1/2}}$$

and the third factor is at most

$$\binom{n}{\lceil n^{1/2} \rceil g_{2k+1}(\lceil n^{1/2} \rceil)},$$

which, by Lemma 1 is subexponential, and in particular is $o(2^{\frac{n}{4k}})$.

Summing now over d, m, p we find that the number of sum-free sets with at least $n^{1/2} g_{2k+1}(n^{1/2})$ elements is less than

$$n^2 2^{\frac{k+1}{2k}(n+2n^{1/2})} 2^{n^{1/2}} \binom{n}{\lceil n^{1/2} \rceil g_{2k+1}(\lceil n^{1/2} \rceil)},$$

and, for n sufficiently large, this is less than

$$2^{\frac{k+2}{2k} n} = o(2^{n(1/2+\varepsilon)})$$

since $k > \frac{1}{\varepsilon}$.

3 The Hausdorff Dimension Of $\mathcal{S}$

The Hausdorff dimension of a set $\mathcal{S}$ contained in the positive integers is defined in the following manner: for two sets $S, T \in \mathbb{N}$, define the distance $d(S,T)$ by $d(S,T) = 2^{-n+1}$ where the sets differ for the first time in the nth position, i. e.

$$\begin{array}{cc} i \in S \Longleftrightarrow i \in T & i = 1, 2, \ldots, n-1 \\ i \in S \cup T, i \notin S \cap T & i = n. \end{array}$$

For any set $\mathcal{T} \subseteq 2^{\mathbb{N}}$ define the **diameter** of $\mathcal{T}$ to be

$$diam(\mathcal{T}) = \sup_{S,T \in \mathcal{T}} d(S,T).$$

For real numbers $\alpha \geq 0$, $\delta > 0$, define

$$\mu_\delta^\alpha(Y) = \inf_{\mathcal{C}} \sum_{C \in \mathcal{C}} (diam(C))^\alpha$$

where the infimum is taken over all countable covers $\mathcal{C}$ of $\mathcal{T}$ satisfying $diam(C) \leq \delta$ for all $C \in \mathcal{C}$ (a cover of $\mathcal{T}$ is a set such that for every $S \in \mathcal{T}$ there exists a set $C \in \mathcal{C}$ such that $S \in C$). Define

$$\mu^\alpha(\mathcal{T}) = \lim_{\delta \to 0} \mu_\delta^\alpha(\mathcal{T}).$$

The Hausdorff dimension of $\mathcal{T}$ is the infimum of those values α for which $\mu^{\alpha'}(\mathcal{T}) = 0$.

As an immediate corollary to Theorem 1 we deduce that the Hausdorff dimension of the set $\mathcal{S}$ of sum-free sets of positive elements is exactly 1/2. Indeed; the dimension is at least 1/2, since the set contains all sets of odd numbers, and this set has Hausdorff dimension 1/2. Further, the dimension is bounded above by

$$\liminf_{n \to \infty} \frac{\log_2 F_{\mathcal{S}}(n)}{n}$$

and since we have

$$\lim_{n \to \infty} \frac{\log_2 F_{\mathcal{S}}(n)}{n} = \frac{1}{2}$$

we see that the dimension of $\mathcal{S}$ is exactly 1/2. This proves a conjecture of Cameron [4].

4 Further Problems

Cameron has also conjectured that the number $F_{\mathcal{S}}(n)$ of sum-free sets is $c2^{n/2}$; it may be possible to prove this by similar techniques, but some sort of additional constraints may be required. Such a result would also imply that the 1/2-dimensional Hausdorff measure of $\mathcal{S}$ is finite; Cameron and the author both believe that this measure is, in fact 1, that is to say, that with respect to this measure, almost every sum-free set consists solely of odd numbers. It seems that this may be an easier problem than that of showing that $F_{\mathcal{S}}(n)$ is $c2^{n/2}$.

References

[1] N.J. Calkin. On the Hausdorff dimension of the set of sum-free sets of positive integers. Unpublished.

[2] N.J. Calkin. *Sum-free Sets and Measure Spaces*. PhD thesis, University of Waterloo, 1988.

[3] P.J. Cameron. Cyclic automorphisms of a countable graph and random sumfree sets. *Graphs Comb.*, 1:129–135, 1985.

[4] P.J. Cameron. Portrait of a typical sum-free set. In C. Whitehead, editor, *Surveys in Combinatorics 1987*, volume 123 of *London Mathematical Society Lecture Notes*, pages 13–42. Cambridge University Press, 1987.

[5] P.J. Cameron. On the structure of a random sum-free set. *Probability Theory and Related Fields*, to appear, 1988.

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