Notes on sum-free and related sets

Peter J. Cameron and Paul Erdős

School of Mathematical Sciences
Queen Mary and Westfield College
Mile End Road
London E1 4NS
U.K.

Mathematical Institute of the Hungarian Academy of Sciences
Reáltanoda u. 13–15
H-1364 Budapest
Hungary

Abstract

Our main topic is the number of subsets of $[1, n]$ which are maximal with respect to some condition such as being sum-free, having no number dividing another, etc. We also investigate some related questions.

In our earlier paper [8], we considered conditions on sets of positive integers (sum-freeness, Sidon sequences, etc.), and attempted to estimate the number of subsets of $[1, n]$ satisfying each condition, and the Hausdorff dimension of the set of subsets of the natural numbers. Most of our effort went into sum-free sets, which have been an obsession of the first author for some time [7].

The present paper is a sequel. Our main concern is to provide similar estimates for the maximal subsets of $[1, n]$ or of the natural numbers satisfying such conditions. We also consider some related questions about sum-free sets: what is the largest sum-free subset of $[1, n]$ with given least element; the existence of large sum-free sets containing no long arithmetic progressions; some classes of sum-free sets defined by irrational numbers; and some remarks on random sum-free sets.

Some of this work was done at the DIMANET workshop held in Mátraháza, Hungary, in October 1995. We thank the organisers for their support, and the participants for their helpful comments.

1 Sum-free sets

1.1 The number of maximal sum-free sets

A subset S of $\{1,\ldots,n\}$ is *sum-free* if $x,y \in S \Rightarrow x+y \notin S$. (We allow $x=y$ here.) The number $f(n)$ of sum-free sets is known to be $2^{(\frac{1}{2}+o(1))n}$. (This bound was found independently by Alon, Calkin, and Erdős and Granville: see, for example, [3]). We have $f(n) \geq c \cdot 2^{n/2}$; we can take any set of odd numbers in $\{1,\ldots,n\}$, or any set of numbers greater than $\frac{1}{2}n$. It is conjectured that $f(n)$ is asymptotically $c \cdot 2^{n/2}$ (Cameron and Erdős [8]).

Since all the sum-free sets just mentioned lie in just two maximal ones, we might expect that the number $f_{\max}(n)$ of maximal sum-free sets is substantially smaller than $f(n)$. So far we have been unable to find a non-trivial upper bound for $f_{\max}(n)$. The following lower bound applies:

Proposition 1.1

$$f_{\max}(n) \geq 2^{\lfloor n/4 \rfloor}.$$

Proof: Let m be either n or $n-1$, whichever is even. Now let S consist of m together with one of each pair of numbers $\{x, m-x\}$ for odd $x < m/2$. Then S is sum-free. It may not be maximal, but no further odd numbers less than m can be added, so distinct sets S lie in distinct maximal sum-free sets.

Problem. Is it true that $f_{\max}(n) = o(f(n))$?

We make a few remarks. In order to show that $f_{\max}(n) = o(2^{n/2})$, it would suffice to prove such a bound for maximal sum-free sets S satisfying $p = \min(S) = o(n)$ and $\frac{1}{10}n \leq |S| \leq \frac{2}{5}n$.

For let $p = \min(S)$. Suppose that $n-p < x \leq n$. If $x = y+z$ for $y,z \in S$, then $y,z \leq n-p$, and $x \notin S$. If not, then x may be added to S without destroying the sum-free property; by maximality, $x \in S$. So S is determined by its intersection with $\{1,\ldots,n-p\}$. Thus the number of maximal sum-free sets with least element p is at most $f(n-p)$, which is $o(2^{n/2})$ if $p > \epsilon n$.

The lower bound on $|S|$ can be assumed because, if $k < \frac{1}{10}n$, then $\binom{n}{k} < 2^{\alpha n}$ with $\alpha < \frac{1}{2}$. The upper bound is deduced from a theorem of Deshouillers, Freiman, Sós and Temkin [9], which asserts that a sum-free set S satisfies one of the following:

(a) S consists of odd numbers;

(b) $|S| \leq \min(S)$;

(c) $|S| \leq \frac{2}{5}(n+2)$.

Clearly there is a unique maximal set satisfying (a). By the previous argument, we may assume that (b) doesn't occur.

The fact that $f_{\max}(n) = o(f(n))$ would follow if a conjecture in our earlier paper were true. We assume that n is even in the following discussion; the results are the same (apart from the constants) for odd n.

We showed that the number $f'(n)$ of sum-free sets with least element greater than $n/3$ is asymptotically $c.2^{n/2}$ for some constant c (which we estimated). Moreover, if $w(n)$ is any function tending to infinity with n, then almost all of these sets have least element greater than $n/2 - w(n)$. We conjectured that almost all sum-free sets either are of this form, or else consist of odd numbers. (It would follow that $f(n) \sim (c+1)2^{n/2}$.)

Consider sum-free sets with least element at least $p = n/2 - w + 1$, where $w < n/6$. Such a set S consists of an arbitrary subset T of $[p, n-p]$, an arbitrary subset U of $[n-p+1, 2p-1]$, and an arbitrary subset of $[2p, n]$ disjoint from $T+T$. Since there are no restrictions on U, we see that if S is maximal then U is the entire interval $[n-p+1, 2p-1]$. So only a proportion $1/2^{3p-n-1}$ of these sets are maximal. If p is chosen so that $p - n/3 \to \infty$, this proportion is $o(1)$.

Of course, there is only one maximal set of odd numbers.

Another approach involves estimating the number of maximal sets containing a given non-maximal set. Since each maximal set has cardinality at least $c\sqrt{n}$ (Theorem 1.2), an affirmative answer to the following question would suffice to give $f_{\max}(n) = o(f(n))$.

Problem. Let S be a subset of $[1, n]$ obtained by deleting one term from a maximal sum-free set. Is it true that S lies in only $o(\sqrt{n})$ maximal sum-free sets?

A related question concerns the Hausdorff dimension of the set $\mathcal{S}_{\max}$ of maximal sum-free subsets of $\mathbb{N}$. It is known that the set $\mathcal{S}$ of all sum-free sets has dimension $\frac{1}{2}$ (see [8]), so $\dim(\mathcal{S}_{\max}) \leq \frac{1}{2}$. The $2^{\lfloor n/4 \rfloor}$ maximal sets constructed above are not relevant to this question, since they cannot be prolonged to subsets of $\mathbb{N}$. However, two related constructions give lower bounds:

(a) The sets consisting of 1 and one of each pair $5k+2, 5k+3$ for $k \geq 1$ lie in distinct maximal sets; so $\dim(\mathcal{S}_{\max}) \geq \frac{1}{5}$.(We require 2^k open balls of diameter $2^{-(5k+3)}$ to cover all these sets.)

(b) We do slightly better by taking the sets constructed by including 2 and then odd numbers with gaps of 4 or 6. Again, these lie in distinct maximal sets. The number x_n of restrictions to $[1, 2n]$ satisfies the recurrence $x_{n+3} = x_n + x_{n+1}$. So $x_n \sim c\alpha^n$, where $\alpha^3 = \alpha + 1$, $\alpha > 1$. Thus $\dim(\mathcal{S}_{\max}) \geq \log\alpha / 2\log 2 = 0.2028\ldots$.

Before leaving this topic, we note the following.

Theorem 1.2 *The size $s^*(n)$ of the smallest maximal sum-free subset of $[1, n]$ satisfies $c_1\sqrt{n} \leq s^*(n) \leq c_2\sqrt{n}$.*

Proof. S is maximal if and only if every element of $[1, n]$ not in S is of the form $x + y$, $x - y$, or $x/2$ for some $x, y \in S$. If $|S| = m$, there are at most $m(m+1)/2 + m(m-1)/2 + m = m(m+1)$ such elements; so $m(m+2) \geq n$, whence $m \geq \sqrt{n+1} - 1$.

The upper bound follows from a construction of Hanson and Seyffarth [12]. They gave a set T of $2t + 1$ residues mod $s = t^2 + 5t + 2$ which is sum-free and has the property that any residue not in T is the sum of two members of T (mod s). It is easily seen that the members of these residue classes in $[1, 2s]$ form a maximal sum-free set.

1.2 Sum-free sets with given least element

We saw in the preceding section that the minimum element of a sum-free set has a significant effect on the structure. Here we consider the function $g(n, p)$, where $g(n, p)$ is the size of the largest sum-free subset of $\{1, \ldots, n\}$ which has smallest element p.

Theorem 1.3

$$g(n,p) \leq \begin{cases} n - p + 1 & \text{if } p > \frac{1}{2}n, \\ p & \text{if } \frac{1}{3}n \leq p < \frac{1}{2}n, \\ \lfloor \frac{1}{2}(n-p) \rfloor + 1 & \text{if } p < \frac{1}{3}n. \end{cases}$$

This bound is attained if either $p \geq \frac{1}{5}(n+2)$ or p is odd.

Proof: If $p > \frac{1}{2}n$, the set $p, p+1, \ldots, n$ contains $n - p + 1$ integers and is sum-free.

Suppose that $\frac{1}{3}n \leq p \leq \frac{1}{2}n$. Suppose that S is sum-free, with $\min(S) = p$. For $2p \leq x \leq n$, S contains at most one of each pair $x - p, x$ of integers; so at least $n - 2p + 1$ integers are excluded, and $|S| \leq (n - p + 1) - (n - 2p + 1) = p$. The bound is realised by the set $S = \{p, \ldots, 2p - 1\}$.

Suppose that $p < \frac{1}{3}n$. Let $q = \lfloor (n-p)/2 \rfloor + 1$, $r = \lfloor (n+p)/2 \rfloor + 1$. Then $p < q$, $p + q = r$, and $q + r - 2 \leq n$. Let $k = |S \cap \{p, \ldots, q-1\}|$ and $l = |S \cap \{q, \ldots, r-1\}|$. Then at least $k + l - 1$ numbers between $p + q = r$ and n are sums of two members of S and so are excluded; so $|S| \leq (n - r + 1) + 1 = \lceil (n-p)/2 \rceil + 1$.

If $n - p$ is even, or if $n - p$ is odd and $n \notin S$, we are finished. If $n - p$ is odd and $n \in S$, then $n - p = (q - 1) + q \notin S$, so we may increase q and r by one in the above argument.

For $\frac{1}{5}(n+2) \leq p \leq \frac{1}{3}n$, there is a sum-free set of size $\lfloor \frac{1}{2}(n-p) \rfloor + 1$, namely $\{p, \ldots, q\} \cup \{2q+1, \ldots, n\}$, where $q = \lfloor \frac{1}{2}(n - p + 1) \rfloor$. For any odd $p \leq \frac{1}{3}n$, there is another set of the same size, namely the odd numbers in $\{p, \ldots, n\}$. This completes the proof of the theorem.

For even $p \leq \frac{1}{5}n$, the results of Deshouillers *et al.* [9] show that $f(n, p) \leq \frac{2}{5}(n+2)$. We do not know if this bound can be attained for $p > 2$. If p is even but not divisible by 5, then the intersection of $\{p, \ldots, n\}$ with either $\{5k+2, 5k+3\}$ or $\{5k+1, 5k+4\}$ has cardinality roughly $\frac{2}{5}(n-p)$. However, for $p > \frac{1}{6}n$, this is beaten by $\{p, \ldots, 2p-1\} \cup \{4p-1, \ldots, 5p-2\}$, with cardinality $2p$.

1.3 The sets S_α

Let α be any irrational real number, and let

$$S_\alpha = \{k \in \mathbb{N} : \tfrac{1}{3} < \{k\alpha\} < \tfrac{2}{3}\},$$

where $\{x\}$ denotes the fractional part of x. The set S_α is sum-free and has density $\frac{1}{3}$. These sets were used by Erdős [11] (see also Alon and Kleitman [2]) to show that any set S of n integers contains a sum-free subset of size at least $n/3$ (namely, its intersection with some $S(\alpha)$). We mention here the open problem of improving this bound to $n/3 + f(n)$, for any function $f(n)$ which tends to infinity.

Calkin and Erdős [5] showed that it is not complete: that is, there are infinitely many numbers not in S_α which have no representation in the form $x+y$ for $x, y \in S_\alpha$. However, S_α is maximal. To see this, let $n \notin S_\alpha$. Suppose that $d = \{n\alpha\} < \frac{1}{3}$ (the other case is similar). Since $\{\{k\alpha\} : k \in \mathbb{N}\}$ is dense in $(0,1)$, there exists $p \in \mathbb{N}$ such that $\frac{1}{3} < \{p\alpha\} < \frac{2}{3} - d$. Then $\frac{1}{3} + d < \{(p+n)\alpha\} < \frac{2}{3}$; so $p, p+n \in S_\alpha$, and we cannot enlarge S_α to a sum-free set by adding n.

This construction may appear to give a large collection of maximal sum-free sets. However, it follows from the next result that $\dim(\{S_\alpha : \alpha \text{ irrational }\}) = 0$.

Proposition 1.4 *Let $S_\alpha(n) = S_\alpha \cap \{1, \ldots, n\}$. The number of sum-free sets of the form $S_\alpha(n)$ for fixed n is*

$$1 + \tfrac{1}{2} \textstyle\sum_{q=1}^{n} \phi(3q) \sim cn^2,$$

where $c = 27/(8\pi^2)$.

Proof: First observe that $S_\alpha = S_{\alpha+m} = S_{m-\alpha}$ for any integer m; so we may assume that $0 < \alpha < \frac{1}{2}$. Now $S_\alpha(n)$ is determined by the least interval containing α whose end-points are $0, \frac{1}{2}$, or of the form $p/(3q)$ where $1 \leq q \leq n$, $1 \leq p < \frac{3}{2}q$, and $3 \not| p$.

We claim that, if α and β lie in distinct intervals of the above form, then $S_\alpha(n) \neq S_\beta(n)$. So suppose that $S_\alpha(n) = S_\beta(n)$. We show by induction that α and β lie in the same interval $(l/3k, m/3k)$. This is true for $k \leq 2$ by inspection. Suppose that it is true for k but false for $k+1$. Assuming that $\alpha < \beta$, it is clear that this can happen only if

$$\begin{aligned} \frac{l}{3k} < \alpha \quad &< \quad \beta < \frac{l+2}{3k}, \\ \alpha < \frac{m}{3(k+1)} \quad &< \quad \frac{m+1}{3(k+1)} < \beta. \end{aligned}$$

These inequalities imply

$$l+\frac{l}{k}<m<l+1+\frac{l+2}{k}.$$

Hence $m=l+1$ or $m=l+2$. But from the form of the intervals, we have $l \equiv 2 \pmod 3$ and $m \equiv 1 \pmod 3$. So $m=l+2>k$. Then it can be checked that the interval $(l-1)/3(k-2), l/3(k-2))$ contains β but not α, a contradiction.

It follows that the number of sets S_α is one more than the number of rationals $p/(3q)$ satisfying the conditions above. This number is $\sum_{q=1}^{n} \phi(3q)/2$, since half of the integers p with $1 \le p \le 3q-1$ and $(p,3q)=1$ are counted.

For the asymptotic estimate, note that $f(n)=\sum_{q=1}^{n}\phi(q)$ is half the number of pairs (p,q) with $1 \le p,q \le n$ and $(p,q)=1$, hence is asymptotically $(3/\pi^2)n^2$. Also, $\phi(3q)=2\phi(q)$ if $3 \not| q$, and $\phi(3q)=3\phi(q)$ if $3|q$. So the sum in question is

$$1+f(n)+\tfrac{1}{2}\textstyle\sum_{q=1}^{\lfloor n/3\rfloor}\phi(q),$$

from which the result follows.

Problem. How many of the sets $S_\alpha(n)$ are maximal?

The sets $S_\alpha(n)$ have another interesting property. There is a tendency for large maximal sum-free sets to contain large arithmetic progressions, of size $\Omega(n)$. (Examples include the odd numbers, the set $\{5k+2, 5k+3 : k \in \mathbb{N}\}$, the set $\{\frac{1}{2}n+1,\ldots,n\}$ for even n, and the set $\{p,\ldots,2p-1\}\cup\{4p-1,\ldots,5p-2\}$). At the Mátraháza meeting, V. Sós asked whether there exist large sum-free sets containing no long arithmetic progression. The largest arithmetic progression contained in $S(\alpha)\cap[1,n]$ depends on the approximability of n. The following observation is based on a remark to the authors by I. Ruzsa. Let $\tau=(1+\sqrt{5})/2$. Recall that $S(\alpha)$ has density $1/3$ for any irrational α.

Theorem 1.5 *The longest arithmetic progression in $S_\tau(n)$ has size $O(\sqrt{n})$.*

Proof. Let $a, a+q, \ldots, a+(m-1)q \in S_\tau(n)$. Then $(m-1)q \le n$, and $\{a\tau\},\ldots,\{(a+(m-1)q)\tau\} \in (\frac{1}{3},\frac{2}{3})$. Hence $\{\pm q\tau\} < 2/(3(m-1))$, and so, for some integer p, we have

$$\left|\tau-\frac{p}{q}\right|<\frac{1}{3(m-1)q}.$$

If $m-1>c\sqrt{n}$, then $q<\sqrt{n}/c$, and so $m-1>c^2q$; so $|\tau-p/q|<1/(3c^2q^2)$. But, if $c>(5/9)^{1/4}$, then $1/(3c^2)<1/\sqrt{5}$, and this inequality has only finitely many solutions.

However, with random methods, we can do better:

Theorem 1.6 *There is a function $c(\epsilon)$ such that, for any $\epsilon>0$, a sum-free set S with density $1/2-\epsilon$ exists for which $S\cap[1,n]$ contains no arithmetic progression longer than $c(\epsilon)\log n$.*

Proof. First we choose a set T with no long arithmetic progressions. Let $p = 1-2\epsilon$, and choose natural numbers independently with probability p. Since the number of m-term arithmetic progressions in $[1,n]$ is at most n^2, the expected number of such progressions in T is at most n^2p^m. If $m > 3\log n/\log(1/p)$, this expected number is less than $1/n$, and so the probability that such a progression occurs is less than $1/n$. Hence the probability that, for some r, the interval $[1,2^r]$ contains an AP of more than $3r\log 2/\log(1/p)$ terms is smaller than $\sum 2^{-r} = 1$. So, with positive probability, this never occurs. Let T be such a set. If $T\cap[1,n]$ contains an AP of length greater than $3\log(2n)/\log(1/p)$, then so does $T\cap[1,2^r]$, where 2^r is the least power of 2 which is not less than n, which is a contradiction since $2^r < 2n$.

To produce a sum-free set, we simply put $S = 2T+1$.

1.4 Random sum-free sets

We next discuss some open problems about random sum-free sets. A random sum-free set S is defined as follows. Consider the natural numbers in turn. If $n = x+y$ with $x,y \in S$, then $n \notin S$; otherwise choose with probability $\frac{1}{2}$ whether $n \in S$ or not (all choices independent).

Let us say that a sum-free set U has *property P* if $\mathrm{Prob}(S \subseteq U) > 0$. Neil Calkin, in his thesis (and see also [4]), showed that property P is equivalent to the condition that, for $n \notin U$, the number of representations $n = x+y$ with $x,y \in U$ grows rapidly (faster than $\log n$) A special case of this, in Cameron [6], is the following. let T be a set of residue classes mod n which is 'complete summ-free', that is, T is sum-free and any residue not in T is the sum of two elements of T (where all calculations are mod n). Examples include the odd numbers ($\{1\}$ mod 2), and the set $\{2,3\}$ mod 5.

Both these sets can be expressed as S_α for *rational* α. For example, we get the odd numbers for $\alpha = \frac{1}{2}$, and the set $\{5k+2, 5k+3 : k \in \mathbb{N}\}$ for $\alpha = \frac{1}{5}$. However, it follows from the result of Calkin and Erdős [5] that S_α does not have property P if α is irrational. Nevertheless, we could ask:

Problem. Is $\mathrm{Prob}(S \subseteq S_\alpha$ for some irrational $\alpha) > 0$?

Analogous constructions of other maximal sum-free sets can be found. The simplest is

$$T_{p,\alpha} = \{k \in \mathbb{N} : \{k\alpha\} \in (p,2p) \cup (1-2p, 1-p)\},$$

where $\frac{1}{8} \le p \le \frac{1}{6}$ and α is irrational. This set has density $2p$. It works because, under these conditions on p, the set $U = (p,2p)\cup(1-2p,1-p)$ is 'complete sum-free mod 1', that is, it is sum-free and any number not in U is the sum of two numbers in U. Many similar sets can be found.

There are many unsolved problems on random sum-free sets. Here are two:

(a) Does a random sum-free set almost surely have a density?

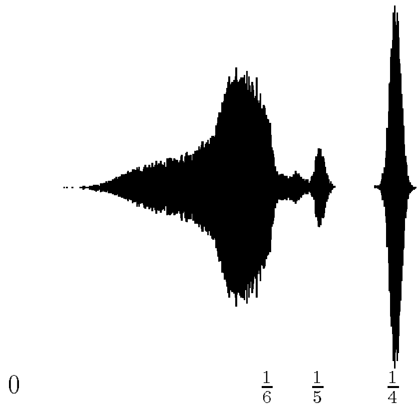

Figure 1: The density spectrum

(b) Is the lower density of a random sum-free set almost surely positive?

Note that an affirmative solution to (b) would imply that a set with property P cannot have density 0.

Empirically, the spectrum of densities seems to be discrete above $\frac{1}{6}$ but to have a continuous part below this value. This could be partly explained if the set of subsets of the sets $T_{p,\alpha}$ above (for all p and α) were shown to have positive measure.

Figure 1 shows an approximation to the density spectrum found by computing many pseudo-random sum-free subsets of $[1, 10000]$.

Another interesting question is the ratio of even to odd members of a random sum-free set. Usually, this ratio appears to be 1 or smaller. (For example, with positive probability, there are no even numbers at all.) However, it tends to a limit greater than 1 with positive probability. For consider the event that all members of S are congruent to a member of T mod $8k + 4$, where

$$T = \{4t + 2 : 0 \leq t \leq 2k, t \neq k\} \cup \{2k + 1, 6k + 3\}.$$

By the results of Cameron [6], this has positive probability, since the set of residue classes is *complete sum-free* (that is, it is sum-free, and every residue class not in the set is the sum of two classes in the set). Moreover, conditioned on this event, the limiting frequencies of the residue classes are equal; so, the ratio of even to odd numbers tends to k.

This suggests several questions. Let $r(n)$ be the ratio of the number of even numbers in $S \cap [1, n]$ to the number of odd numbers in this set. Does the event $\limsup r(n) = \infty$ have positive probability? Which numbers r have the property that $\limsup r(n) = r$ has positive probability? Does $r(n)$ almost surely tend to a limit?

2 Maximal sets in other families

In our earlier paper, we estimated the number of subsets of $[1, n]$ satisfying various conditions, of which being sum-free was one. In each of these cases, one can ask for similar estimates for the number of maximal subsets satisfying the appropriate condition. We now give results on some of these cases.

Following our earlier notation, we let $\mathcal{A}$ be a class of sets of integers closed under taking subsets; $s(n)$ denotes the size of the largest member of $\mathcal{A}$ contained in $[1, n]$, and $f(n)$ the number of sets in $\mathcal{A}$ contained in $[1, n]$ (so that $2^{s(n)} \leq f(n) \leq n^{s(n)}$). Also, we let $f_{\max}(n)$ be the number of sets in $\mathcal{A}$ which are maximal with respect to being contained in $[1, n]$. Another interesting parameter is $s^*(n)$, the size of the smallest such maximal set.

We note that the method of choosing a random sum-free set described earlier can be adapted to the situations described below. We hope to investigate this further subsequently.

2.1 A general result

Theorem 2.1 *Let $\mathcal{A}$ be closed under taking subsets and also closed under translation, and suppose that the size $s(n)$ of the largest member of $\mathcal{A}$ contained in $[1, n]$ satisfies $s(n) = o(n)$. Then $\limsup f_{\max}(n) = \infty$.*

Remark. This theorem applies to the set of *Sidon sequences* (having all pairwise sums distinct), and to the class of sequences containing no k-term arithmetic progression for fixed k. (For Sidon sequences, $s(n) = O(\sqrt{n})$. The fact that $s(n) = o(n)$ in the second case is Szemerédi's Theorem [14].)

Proof. In our previous paper, we showed under the same hypotheses that $\limsup f(n)/2^{s(n)} = \infty$. Now the result follows, since $f(n) \leq f_{\max}(n) \cdot 2^{s(n)}$.

2.2 All partial sums distinct

Let $\mathcal{A}$ be the set of sequences $a_1 < a_2 < \ldots < a_t \leq n$ for which all subsums $\sum \epsilon_i a_i$ (with $\epsilon_i = 0$ or 1) are distinct. In [8] we showed that

$$n^{(1+o(1))\log n/\log 3} \leq f(n) \leq n^{(1+o(1))\log n/\log 2}.$$

Theorem 2.2

$$f_{\max}(n) \geq f(n)/cn \log n.$$

Proof. Let such a maximal sequence S have t terms. Since the 2^t partial sums are distinct,

$$2^t - 1 \le tn \le cn \log n.$$

But $2^t - 1$ is the number of non-empty subsequences of S. The result follows.

2.3 Requiring divisibility

Let $\mathcal{A}$ be the family of sequences $a_1 < a_2 < \ldots$ for which a_i divides a_j whenever $i < j$. In [8], we observed that the number $f(n)$ of such sequences in $[1, n]$ satisfies $f(n) \sim cn^{\alpha}$, where $\alpha = 1.7286\ldots$ is the root of $\zeta(\alpha) = 2$.

Maximality of such a sequence means that $a_1 = 1$ and all the factors a_{i+1}/a_i are prime. Hence, the number $g(n)$ of maximal sequences with largest term n is given by the recurrence

$$g(1) = 1, \qquad g(n) = \sum_{\substack{p \text{ prime} \\ p|n}} g(n/p);$$

and we have

$$f_{\max}(n) = \sum_{n/2 < m \le n} g(m).$$

So certainly $f_{\max}(n) \ge n/2$.

In fact, $f_{\max}(n)$ grows faster than linearly. For let $n = (2.3.5.7.11.13)^k$. Then $f(n)$ is at least the number of orderings of the factors of n, which is $(6k)!/(k!)^6 > (6^6 - \epsilon)^k$ for large k. Thus,

$$\limsup \log f_{\max}(n)/\log n \ge \log(6^6)/\log(2.3.5.7.11.13) = 1.042736\ldots$$

(Since $\log(p_1 \cdots p_r) \sim \log(r^r)$, where $p_1, \ldots, p_r$ are the first r primes, we cannot do much better by this method.)

Problem. Is it true that $f_{\max}(n) \sim dn^{\beta}$ for some constants d and β?

2.4 Forbidding divisibility

Let $\mathcal{A}$ be the family of sequences where no term divides another. In our earlier paper, we showed that $c_1^n \le f(n) \le c_2^n$, where $c_1 = 1.55\ldots$ and $c_2 = 1.59\ldots$. We conjectured, but were unable to prove, that $f(n)^{1/n}$ tends to a limit as $n \to \infty$.

It is clear that $f_{\max}(n) = o(f(n))$. For any maximal subset of $[1, n]$ in which no term divides another must contain all the primes in the interval $(n/2, n]$, so that

$$f_{\max}(n) \le f(n)/2^{\pi(n) - \pi(n/2)}.$$

We cannot yet show that $f_{\max}(n)$ is exponentially smaller than $f(n)$. Some remarks on this follow.

First, the maximum size $s(n)$ of such a set is $\lceil n/2 \rceil$. For, given any odd number $m \leq n$, at most one number of the form $2^a \cdot m$ can belong to the set. If the bound is attained, then the set contains one number of this form for each value of a. Such a set is described by a function $a(m)$ from the odd numbers in $[1, n]$ to the non-negative integers, satisfying

(a) $a(m) \leq \log(n/m)/\log 2$;

(b) if m_1 divides m_2, then $a(m_1) > a(m_2)$.

The number of functions satisfying (a) is

$$\prod r^{\lceil n/2^{r+1} \rceil} \sim c^n,$$

where $c = \prod r^{1/2^{r+1}} = 1.2891\ldots$. This is exponentially smaller than our lower bound for $f(n)$.

There is also an exponential lower bound for the number of such subsets of cardinality $\lceil n/2 \rceil$. For each integer x in $(n/3, n/2]$, include either x or $2x$ in the set; then include all integers in $(n/2, 2n/3]$, and all odd integers in $(2n/3, n]$. This gives about c^n sets, where $c = 2^{1/6} = 1.12246\ldots$.

We also have some information on the small maximal sets. To avoid trivialities, we do not allow the number 1.

Theorem 2.3 *The size $s^*(n)$ of the smallest maximal set in $\mathcal{A}$ contained in $[2, n]$ is $\pi(n)$. There are at least $2^{cn/\log n}$ maximal sets of this size.*

Proof. Let $a_1 < a_2 < \ldots < a_k \leq n$ be such a maximal set. We prove that, given any set P of m primes in $[1, n]$, there are at least m terms of the sequence divisible by some prime in P. The proof is by induction on m. For $m = 1$, if there were a prime dividing no term of the sequence, it could be added to the sequence.

Suppose that the result is true for $m - 1$ but false for m. Let P be a set of m primes whose members divide fewer than m terms of the sequence. Let p be the smallest prime in P, and p^k the largest power of p dividing any term in the sequence. It cannot occur that a term in the sequence is divisible by p^k and no other prime in P, else the remaining $m - 1$ primes would divide fewer than $m - 1$ terms. So there is some term $p^k qx$ in the sequence, where q is a prime greater than p. Then $p^{k+1} < p^k qx \leq n$, and we may add p^{k+1} to the sequence. This completes the induction.

The proof shows that the conditions of Hall's Theorem hold; so, if $p_1 < p_2 < \ldots < p_r$ are the primes in $[1, n]$, with $r = \pi(n)$, then we can number the sequence as $a_1, a_2, \ldots, a_r$ so that p_i divides a_i for all i. The argument shows that if a set of m primes divide only m terms of the sequence, and p_i is the smallest of the primes, then a_i is divisible by no other prime in the set. Applying this inductively to $p_i, p_{i+1}, \ldots, p_r$, we see that, if p_j divides a_i, then $j \leq i$.

We construct many maximal sequences as follows. Start with the set of primes in $[1,n]$. Let p be any prime less than $\sqrt{n}$. For each prime q satisfying $\sqrt{n} < q \le n/p$, multiply q by any power of p (so that the result is still less than n). Then replace p by any power of itself strictly greater than the powers used in the first part of the construction. The result is still in $\mathcal{A}$ and is still maximal.

If we choose $p=2$, and replace some primes $q \in (\sqrt{n}, n/2]$ by $2q$ and then replace 2 by 4, we obtain $2^{\pi(n/2)-\pi(\sqrt{n})}$ maximal sets of size $\pi(n)$. This is of order $2^{cn/\log n}$ as claimed: the more general construction merely improves the value of c slightly.

Problems. How many maximal sets are there? Does $f_{\max}(n)^{1/n}$ tend to a limit? How many maximal sets of size $\pi(n)$?

2.5 Any two terms coprime

Let $\mathcal{A}$ be the set of sequences with any two terms coprime. Apart from 1, the terms of such a sequence have disjoint sets of prime factors, so there are at most $\pi(n)$ of them. We proved in [8] that the number of sequences is at least $\mathrm{e}^{(1/2+o(1))\sqrt{n}}2^{\pi(n)}$ sequences; so there are at least $\mathrm{e}^{(1/2+o(1))\sqrt{n}}$ maximal ones.

Also, there are at most $\mathrm{e}^{(2+o(1))\sqrt{n}}$ maximal sequences. For such a sequence contains at most $t=\pi(\sqrt{n})$ non-prime terms, which can be chosen in at most

$$\sum_{i=0}^{t}\binom{n}{i} = \mathrm{e}^{(2+o(1))\sqrt{n}}$$

ways; then we must include all primes not dividing these chosen terms.

2.6 No two terms coprime

Let $\mathcal{A}$ be the set of sequences of integers with no two terms coprime. We proved in [8] that $f(n) \le n \cdot 2^{\lfloor n/2 \rfloor}$. On the other hand, $f(n) \ge 2^{\lfloor n/2 \rfloor}$, since any set of even numbers belongs to $\mathcal{A}$. This construction gives only one maximal sequence. However, we show that there are relatively many of them:

Theorem 2.4

$$f_{\max}(n) \ge 2^{n^{c/\log\log n}}.$$

Proof. We deduce the bound from the following construction. An *antichain* in $\mathcal{P}[1,m]$ is a family of subsets of $[1,m]$ no one of which contains another. The *dual*, or *blocker*, of the antichain $\mathcal{X}$ is the antichain $\mathcal{X}^*$ consisting of all sets which are minimal with respect to intersecting every set in $\mathcal{X}$. It is known that $\mathcal{X}^{**}=\mathcal{X}$ (Edmonds and Fulkerson [10]; see Seymour [13] for a short proof). We call $\mathcal{X}$ *self-dual* if $\mathcal{X}^*=\mathcal{X}$.

Theorem 2.5 *There are at least*

$$2^{\binom{2k-1}{k-1}-1} / \binom{2k-1}{k-1}^{\binom{2k}{k-1}/2^k} = 2^{(2-o(1))^m}$$

self-dual antichains in $\mathcal{P}([1,m])$*, for* $m = 2k$*.*

Proof. We call an antichain $\mathcal{X}$ *sub-dual* if $\mathcal{X} \subseteq \mathcal{X}^*$. Now $\mathcal{X}$ is sub-dual if and only if it has the following properties:

(a) $\mathcal{X}$ is intersecting;

(b) for any $X \in \mathcal{X}$ and $x \in X$, there exists $y \in \mathcal{X}$ such that $X \cap Y = \{x\}$.

If this holds, and Y is any member of $\mathcal{X}^* \setminus \mathcal{X}$, then for each $y \in Y$ there exists $X \in \mathcal{X}$ with $X \cap Y = \{y\}$ (else $Y \setminus \{y\}$ would meet every set in $\mathcal{X}$, and Y would not be minimal). Hence $\mathcal{X} \cup \{Y\}$ is also sub-dual. It follows that any sub-dual antichain can be enlarged to a self-dual antichain.

Our strategy is as follows. If $m = 2k$, any choice of one from each complementary pair of k-subsets of $[1,m]$ gives an intersecting antichain. We show that, for many choices, condition (b) is also satisfied, so that we have many sub-dual antichains. These can be enlarged to self-dual antichains; but in this enlargement process, no further k-sets will be added, so that the resulting self-dual antichains are all distinct.

A choice C of one of each complementary pair of k-sets can be regarded as a vertex of the hypercube of dimension $\binom{2k}{k}/2 = \binom{2k-1}{k-1} = N$, say. So there are are 2^N choices altogether. Now the failure of condition (b) means that there is a $(k-1)$-subset A such that every k-set containing it is in C. (Take $A = X \setminus \{x\}$, with x and X witnessing a failure of (b).) Call such a subset *bad*. Now, in a random choice, the probability that a given set A is bad is $1/2^{k+1}$; so the expected number of bad sets is $\binom{2k}{k-1}/2^{k-1} = \mu$, say. It follows that, for at least half of all possible choices, the number of bad sets is not greater than 2μ.

Suppose that A is bad (for a choice C), so that every k-set containing A is in C. Then any $(k-1)$-set disjoint from A lies in at least two k-sets not in C. Hence, if we replace any k-set containing A by its complement, we obtain a new choice C' in which A is no longer bad, and no new bad set is created. Note that C and C' are adjacent in the N-dimensional hypercube. It follows that a choice with r bad sets lies at distance r in the hypercube from a choice with no bad sets (a *good choice*, we will say). So the balls of radius 2μ centered on the good choices cover at least half the vertices of the hypercube.

The size of such a ball is not greater than $N^{2\mu}$. So there are at least $2^{N-1}/N^{2\mu}$ good choices, as claimed.

Proof of Theorem 2.4. Choose m such that $p_1 p_1 \cdots p_m \le n$, where $p_1, p_2, \ldots, p_m$ are the first m primes. Let $\mathcal{X}$ be a self-dual antichain on $[1,m]$. For each $X \in \mathcal{X}$, let

$$p(X) = \prod_{i \in X} p_i.$$

Let A be the set of numbers in $[1,n]$ divisible by $p(X)$ for some $X \in \mathcal{X}$. Since $\mathcal{X}$ is intersecting, no two members of A are coprime. If r has a common factor with each number in A, then the set $Y = \{i \in [1,m] : p_i | r\}$ meets every set in $\mathcal{X}$. By the self-duality, Y contains some set $X \in \mathcal{X}$, whence $p(X)|r$, and $r \in A$. So A is maximal.

Now $m^m \sim n$, so $m \geq c \log n / \log \log n$. Thus the number of maximal sets in $\mathcal{A}$ is at least

$$2^{(2-o(1))^m} \geq 2^{n^{c/\log\log n}}.$$

References

[1] R. Ahlswede and L. H. Khachatrian, On extremal sets without coprimes, *Acta Arith.* **66** (1994), 89–99.

[2] N. Alon and D. J. Kleitman, Sum-free subsets, pp. 13–26 in *A tribute to Paul Erdős* (ed. A. Baker, B. Bollobás and A. Hajnal), Cambridge University Press, Cambridge, 1990.

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

[4] N. J. Calkin, On the structure of a random sum-free set, to appear.

[5] N. J. Calkin and P. Erdős, On a class of aperiodic sets, *Math. Proc. Cambridge Philos. Soc.* **120** (1996), 1–5.

[6] P. J. Cameron, On the structure of a random sum-free set, *Probab. Theory Rel. Fields* **76** (1987), 523–531.

[7] P. J. Cameron, Portrait of a typical sum-free set, *Surveys in Combinatorics 1987* (ed. C. Whitehead), 13–42, *London Math. Soc. Lecture Notes* **123**, Cambridge University Press, Cambridge, 1987.

[8] P. J. Cameron and P. Erdős, On the number of sets of integers with various properties, *Number Theory* (ed. R. A. Mollin), 61–79, Walter de Gruyter, Berlin, 1990.

[9] J.-H. Deshouillers, G. Freiman, V. Sós and M. Temkin, On the structure of sum-free sets, to appear.

[10] J. Edmonds and D. R. Fulkerson, Bottleneck extrema, *J. Combinatorial Theory* (A) **8** (1970), 299–306.

[11] P. Erdős, Extremal problems in number theory, *Proc. Symp. Pure Math.* **8** (1965), 181–189.

[12] D. Hanson and K. Seyffarth, K-saturated graphs of prescribed minimal degree, *Congressus Numerantium* **42** (1984), 169–182.

[13] P. D. Seymour, The forbidden minors of binary clutters, *J. London Math. Soc.* (2) **12** (1976), 356–360.

[14] E. Szemerédi, On sets of integers containing no k elements in arithmetic progression, *Acta Arith.* **27** (1975), 299–345.